HERIOT-WATT UNIVERSITY, EDINBURGH

DEPARTMENT OF CIVIL ENGINEERING

THE RELIABILITY ANALYSIS OF GEOTECHNICAL STRUCTURES

by

Geoffrey N. Smith,

D.R.T.C., M.Sc., C.Eng., M.I.C.E.

Submitted for the degree of Doctor of Philosophy.

July, 1984

ABSTRACT

For some years now the author has concerned himself with the possibility of producing a new, yet simple method, by which the reliability analysis of geotechnical structures could be carried out.

Such a method has now been devised and is detailed in this thesis along with worked examples of a practical nature which illustrate how it is used.

A "state of the art " presentation of those aspects of statistics and probability theory that can be of assistance to the practising consultant civil engineer is also given.

ACKNOWLEDGEMENTS

The text of this thesis represents part of the work carried out by the writer under a research consultancy with the Transport and Road Research Laboratory, Crowthorne, Berkshire and he would like to express his thanks for the support that he has received from various people within that establishment.

In particular the author would like to express his gratitude to Mr. J.B. Boden, now Head of the Geotechnical Section of the Building Research Establishment at Garston, Hertfordshire, and to Dr. R.T. Murray, Deputy Head of the Ground Engineering Divison, T.R.R.L. at Crowthorne, who inititated the research.

The author's thanks also go to Mr. M.P.O'Reilly, for his interest and encouragement in the work since being appointed Head of the Ground Engineering Division at T.R.R.L. in 1980.

The writer also wishes to acknowledge the help and encouragement that he has received from Heriot-Watt University and he particularly wishes to thank Professor A. Bolton, of the Department of Civil Engineering, for his friendly advice during the period of the work.

LIST OF CONTENTS

INTRODUCTION

Until recently the suitability of a structural element to withstand a particular loading, or to not deflect more than a prescribed limit, was measured by means of a single number, its factor of safety, F.

The factor of safety method employs clear logic and involves straightforward mathematics. If, for instance, part of a structural design involves the ultimate limit state analysis of a reinforced concrete beam then the factor of safety of the beam can be found from the expression:-

$$F = \frac{\text{Maximum bending moment}}{\text{Allowable bending moment}}$$

In terms of stresses the expression can be written as:-

$$F = \frac{\text{Maximum tensile stress induced by bending}}{\text{Allowable tensile bending stress}}$$

The philosophy behind the factor of safety approach is simply that, provided F is equal to or greater than a predetermined value, obtained from the relevant code of practice, then the beam is considered safe.

One drawback is that there is always the temptation for the designer, in order to achieve complete safety, to uneconomically strengthen the beam in order to achieve a high value for F.

Uncertainty in structural engineering

Few would argue that complete safety is ever possible. Any design method that might be used for the beam must involve uncertainties.

Structural uncertainties are well known and well documented so that there is little need here for more than a brief mention of the main ones.

The magnitudes of the applied forces will largely be indeterminate.

The strengths of the structural materials involved will vary with both the material and the quality control used in its manufacture or selection.

Structural dimensions, particularly the depths to the reinforcement, can be significantly different to the dimensions shown on the drawings.

Uncertainty in geotechnical engineering

When one considers a geotechnical structure, such as an embankment or a retaining wall, a whole new range of uncertainities are added to the ones encountered in structural engineering not the least being that the designer must work with materials whose properties have not been specified but have been provided by nature.

The major uncertainties connected with geotechnical engineering are set out and briefly described below.

1. Spatial variablity

The subsurface soil may consist of a set of strata of different materials each of which should be considered as a discrete layer.

Another form of spatial variability is when an apparently homogeneous soil has material properties which vary in value from point to point, either randomly or in some form of pattern, or both.

2. Limited number of test results

Statistical uncertainty is often caused by the economical necessity of keeping the number of samples, in-situ and laboratory, to a minimum. Measured values will vary between samples not only because of spatial variability but also because of errors introduced during the tests. In theory these uncertainties can be reduced, to a required level, by an appropriate increase in the number of samples tested.

3. Bias errors

In most geotechnical tests there is a systematic difference between the measured and the actual value of a particular parameter, Ladd (1977).

4. Model uncertainty

Errors are induced by numerical methods, simplifications in seepage problems, soil stress-strain behaviour being greatly simplified so that a model able to arrive at predictions can be created, etc. It is generally agreed that model uncertainty in geotechnics is usually large.

5. Applied loading

The magnitude and distribution of the applied loads will be uncertain.

6. Omissions

No matter how comprehensive a design analysis there will always be something left out, mainly due to lack of knowledge.

The probability of failure

Perhaps not suprisingly there are many recorded examples of geotechnical structures where the calculated factor of safety exceeded 1.0 and yet failure occurred. F, being a number obtained by a deterministic method, cannot allow for the variability of the soil parameters involved.

Lumb (1970) summed up the situation:-

"The traditional safety factor concept has the serious disadvantage that the actual variablity of the soil strength is not directly taken into account and, consquently, a particular conventional safety factor value does not necessarily have the same meaning for all soils. Comparison of different designs with different soil types, or even different designs with the same soil type, is not easy, unless the conventional safety factors are so large as to preclude any practical risk of failure."

For a geotechnical structure the factor of safety is really a random variable whose variability is due exclusively to the variability of the applied loads and the soil parameters involved.

If failure is defined as the event of F achieving a value equal to or less than 1.0 then the probability of this event is equal to the probability of failure ,P_f.

$$P_f = P[\text{Failure}] = P[F \leq 1]$$

Formal treatment of design uncertainties

It must be noted that not all uncertainties are capable of formal treatment, de Mello (1977). Extreme events such as seismic activity, internal erosion in an earth dam, accidental damage, explosions, etc. are difficult to model and are usually more realistically dealt with by rough empirical guides and the adoption of a conservative design.

Traditionally geotechnical uncertainties have often been allowed for by the use of a suitable observation and modification programme carried out during construction, Peck, (1962), Casagrande, (1965).

However in many geotechnical problems average values, rather than extreme values, are dealt with. It is possible to design models for such situations and then to attempt to minimise the uncertainties with the aid of statistics and probability theory.

A suggested method for evaluating P_f for geotechnical structures is described in this thesis together with a brief description of the statistics and probability theory involved.

CHAPTER ONE - BASIC PROBABILITY THEORY

Sets

The study of events and the probability of their happenings inevitably draws one towards the idea of the set.

In a test series of measurements the mean value obtained is an event resulting from the whole set of measured values.

A set is therefore a collection of items and, as with an event, is usually designated by a capital letter, A, B, C, etc. The individual elements that make up a set are generally denoted by lower case letters, a, b, c,....

Eg. Set A = a_1, a_2, a_3, a_4

= a_3, a_4, a_1, a_2 (as arrangement of elements does not affect a set)

The convention $a_1 \in A$ simply means that a_1 is an element of the set A.

In most civil engineering situations a set is defined by the listing of the elements within it, such as the measurements obtained for a particular test. However there are often occasions when it is not possible to determine the total elements of a set, although we know they exist, such as the infinite set of soil samples that could be collected from a particular stratum.

In such a situation, although the full set cannot be listed, the properties of the set can. For example a set B, consisting of all even numbers between 2 and 100, could be specified as:-

$B = \{b \; ; \; b \text{ is an even number between 2 and 100}\}$

where ";" means "given that" or "such that".

Obviously set B could also have been listed as:-

$B = \{2, 4, 6, 8, \ldots\ldots 98, 100\}$

The universal set

The complete collection of all possible elements of a set is known as the universal set or the sample space and given the symbol Ω, the Greek letter omega, or the capital letter S.

Fig. 1.1 shows the sample space, i.e. all the possible events involved in the scores obtained from the throwing of two dice.

The sample space, such as the total 36 elements of Fig.1.1, represents the certain event, in this case the event "there will be some score".

An impossible event is one which is outwith the sample space, such as the event (7,1) in Fig.1.1.

The subset

If B is a set of elements taken from a universal set A then B is referred to as a subset of A. This is expressed as:-

$$B \subset A \quad \text{or} \quad A \supset B$$

meaning "B is contained in A" or "A contains B" respectively.

In Fig.1.1 the subset {(4,1),(3,2),(2,3),(1,4)} represents the total number of ways of obtaining the score "5". The event "scoring 5" is by no means a single event as it can occur in 4 different ways. An event that can occur in more than one way is called a compound event whereas a single event would be the scoring of double 1 as there is only one element in the sample space that represents this event,(1,1).

Union of sets $(A \cup B)$

The union of two sets, A and B, is the set which contains all the elements common to either A or B.

Intersection of sets $(A \cap B)$

The intersection of two sets, A and B, is the set which contains all the elements that are in both A and B.

The Venn Diagram

A sample space, or universe, S, and its subsets can be presented in a pictorial form by the use of this technique.

The universal set, S, is represented as a rectangle with its subsets lying within it. (Fig.1.2).

Difference between sets

If B is a subset of A then A > B and the set (A-B) is called the complement of B relative to A and given the symbol $\bar{B}_A$.

If S is the sample space then the set (S - B) is known as the the complement of B and given the symbol $\bar{B}$.

If $b \notin B$ then b is not a member of B and must therefore be an element in $\bar{B}$, the complement set of B.

If there are two sets A and B, then the complement of $A \cup B$ is denoted as $\overline{A \cup B}$.

The Venn diagrams of Fig.1.3 illustrate various set operations.

The algebra of sets and events

An event is a subset of the set of total possible events so the algebra applicable to sets is identical to that for events.

The most important theorms of set algebra are set out below and can be demonstrated by a study of the appropriate Venn diagrams.

Commutative law : $A \cup B = B \cup A$

$A \cap B = B \cap A$

Associative law : $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$

$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$

Distributive law : $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Complementary Laws : $A - B = A \cap \overline{B}$

De Morgan's Laws) $\overline{A} \cup \overline{B} = \overline{A \cap B}$

$\overline{A} \cap \overline{B} = \overline{A \cup B}$

Example 1.1

By means of Venn diagrams prove the theorem $A - B = A \cap \overline{B}$

Solution

When stated in words the theorem is, "The elements contained in a set A but not in a set B are the same elements common to both set A and the complement of set B".

If set A and set B are as shown in Fig.1.4A then the difference set A - B is represented by the hatched area shown.

The dotted area of Fig.1.4B represents the complement set $\overline{B}$. and it is fairly obvious that the dotted area of Fig.1.4C, which represents the elements common to A and $\overline{B}$, is the same as the hatched area of Fig.1.4A.

Hence $A - B = A \cap \overline{B}$

Note

The above axiom can be illustrated by considering the elements within the sets:-

$A - B = [x \; ; \; x \in A \text{ and } x \notin B]$

$$= [x \; ; \; x \in A \text{ and } x \in \bar{B}] = A \cap \bar{B}$$

Probability

The probability that a particular event, A, will happen is expressed mathematically as P[A].

If the event A will never happen, Eg. pigs will fly, then the value of P[A] will be 0 whereas if event A will happen, Eg. the world will end sometime, then P[A] = 1.

Probability values are classified in one of two ways, depending upon how they were estimated.

A probability value determined with no prior knowledge, i.e. with a priori judgement, is called a "prior probability value".

A probability value estimated with the use of relevant information drawn from previous experience involves a posteriori judgement and is called a "posterior probability value".

For most civil engineering design work posterior probabilities are generally found by some method based on a frequenistic approach. For example if, after N number of tests, an event A occured n times then it is said that the probability of A happening in any future test is n/N.

Whilst suitable for most design situations the frequenistic approach cannot be applied to the case of an unrepeatable event, which is usually the case when making a design decision.

Any prior knowledge used for the estimation of such a posterior probability value can only be obtained from subjective judgement, based on previous experience, and the resulting probability value is really a "degree of belief" posterior probability value.

Many engineers, familiar with the frequenistic approach, experience difficulty in accepting this idea of degree of belief.

For instance, few would be willing to accept the idea that the probability of a rock fault existing at some site is 60%.

Most would argue that as the fault either exists or does not exist then the probability is either 1 or 0.

It is in these situations that Baye´s Theorem, described later in this chapter, can be of assistance.

With an a priori approach the statements P[A] =1 and P[A] = 0 imply absolute certainty.

However, with posteriori judgement, one cannot assume absolutely that because an event happened in the past it will do so again in the future.

Similarly, with the degree of belief approach, the statement that P[A] = 1 means that it is considered that A will occur, not that it must occur.

Example 1.2 - Prior probability

The probability of drawing an ace from a full pack of cards.

There are 4 aces in a pack and a total of 52 cards.

Hence n = 4 and N = 52

Probability of drawing an ace = 4/52 = 1/13

Example 1.3 - Posterior probability

45 control tests were carried out on a long stretch of compacted subgrade. 5 tests yielded results below specification.

If a further ten tests had been carried out how many of these tests could be expected to have given results below specification?

Probability of results below specification = 5/45 = 1/9

For a further set of ten tests expected number of tests below specification = 10x1/9 = 1.1, i.e. 1 test.

Mutually exclusive events

If there is a set of events A, B, C,..... such that the happening of one excludes the happening of the others then we say that A, B, C,.... are mutually exclusive.

An example of mutual exclusion would be the acceptance of a tender from among several submitted. If contractor A is successful in his bid then there is no possibility of contactors B, C, etc. also being successful.

The summation law - union probability

This law applies to mutually exclusive events and states that for a series of mutually exclusive events, the union probability of at at least one of these events occurring is equal to the sum of the separate probabilities of the events.

Consider three events, A, B and C. The probability that any one of these events will occur is:-

$$P[A \cup B \cup C] = P[A] + P[B] + P[C]$$

Note the use of the symbol $\cup$ to represent "or".

Example 1.4

Examples of the summation law are:-

i) With the tossing of a fair coin:

The probability of a head = P[A] = 0.5 or 50%

The probability of a tail = P[B] = 0.5 or 50%

Probability of either a head or a tail =

P[A] + P[B] = 1.0 or 100%.

ii) With a set of strength measurements of a particular material:

P[A] = The probability of the actual strength being equal to or less than the mean value = 0.5.

P[B] = The probability of the actual strength being equal to or greater than the mean value = 0.5.

P[A] + P[B] = The probability that the actual strength is either equal to or is greater or smaller than the mean value = 1.0.

Independent events

If we have a set of possible events such that the happening of any one event has no effect on the probabilities of the happening of the others then the events are said to be independent.

If a perfect random number generator is programmed to produce integers between 1 to 100 then the production of each number by the generator will be an independent event.

This means that if the generator was to produce the figure 24 in two consecutive intervals then the chance of it producing a further 24 in the next interval is exactly the same as its chance of producing any of the other numbers.

The multiplication law - joint probability

This law applies to independent events and states that for a series of independent events, the joint probability of all of the events occurring is equal to the product of the separate probabilities of the events.

In terms of three events, A, B and C, the law can be expressed as:-

$$P[A \cap B \cap C] = P[A] x P[B] x P[C]$$

Note the use of the symbol $\cap$ to represent "and".

Other conventions used to write probability expressions are:-

i) $P[A \cap B \cap C]$ is often written as P[ABC]

ii) P[A]xP[B]xP[C] usually written as P[A]P[B]P[C].

Whenever there is no risk of ambiguity these later conventions will be used in the text.

Example 1.5

Probability independence is illustrated by the tossing of dice:-

If two dice are thrown what is the probability of two threes?

Let P[A] = probability of a three on the first die (= 1/6)

Let P[B] = probability of a three on the second die (= 1/6)

Then probability of two threes, $P[A B] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

A further look at union probability

Considering the previous example. What is the union probability of either A or B? (i.e. the probability of obtaining a three on either die or on both?

If we use the summation law in the form stated above then

$$P[A \cup B] = P[A] + P[B] = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

However, if we obtain this probability by enumeration we achieve a different value.

The set of events that cause a 3 to be scored on either die is a subset of the sample space shown in Fig.1.1 and is:-

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(1,3) (2,3) (3,3) (4,3) (5,3) (6,3)

A total of 12 events out of a total of 36 which gives a probability value of 12/36 equalling the 1/3 value obtained from the formula.

However, if the subset of the 12 events is examined we see that the event (3,3), the probability of a 3 being scored on each die, has been included twice.

The formula is at fault as, in this problem, the events A and B are independent and the joint probability that they may occur together, $P[A \cap B]$, has been included twice.

Things can be put right by simply subtracting $P[A \cap B]$ from the value obtained for $P[A \cup B]$ to give :-

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = \frac{1}{3} - \frac{1}{36} = \frac{11}{36}$$

This is the general form of the summation law and applies to all events whether mutually exclusive or not. (If A and B are mutually exclusive then $P[A \cap B] = 0$).

The complement of a probability

Often when considering the union of several probabilities the numerical work can be reduced if we think in terms of the complements of these probabilities.

For any event A :-

$$0 \leq P[A] \leq 1$$

This means that in probability analysis the total probability space, S, equals unity and $\overline{P[A]} = 1 - P[A]$.
i.e. For any event A the complement of its probability is equal to one minus its probability.

A convenient use for the complement of a probability is when it is required to estimate the probability of occurrence of a single event over a given number of trials.

Assume that, in one trial, the probability of occurrence of an event A is P[A].

Then the probability of non-occurrence of A = $1 - P[A]$

And the probability of occurrence of A in n trials = $1 - (1-P[A])^n$

Reliability

An important probability complement is the reliability. If a system has a probability of failure of 10% then the system is 90% reliable.

$$\text{i.e. Reliability, } R, = 1 - P_f$$

Example 1.6

In a certain region the subgrade is predominantly a silty soil with the odd clay lens. The average size of these lenses is $750m^2$.

If on a site $100 \times 150m^2$ a clay lens exists what is the probability of encountering it in any one of 8 randomly placed boreholes?

Solution

The probability of finding the lens in one borehole, P[L], can be estimated from the ratio of the two areas:-

$P[L] = \frac{750}{100 \times 150} = 0.05 \quad \Rightarrow \overline{P[L]} = 1 - P[L] = 0.95$

Probability of encountering the lens in at least one of the eight boreholes, P[F], can now be found:-

$$P[F] = 1 - 0.95^8 = 0.337 \text{ .... say } 33\%$$

Example 1.7

In a reliability analysis for a proposed concrete retaining wall the following probabilities of failure were obtained.

Risk of bearing capacity failure $= P_b = 0.03$

Risk of overturning $= P_o = 0.01$

Risk of sliding failure $= P_s = 0.02$

Risk of structural concrete failure $= P_c = 0.03$

Determine a value for P_f, the probability of failure of the wall.

Solution

Assuming that the various modes of failure are independent it is obvious that the occurrence of anyone of these failure events will result in the failure of the wall.

$$P_f = P_b \cup P_o \cup P_s \cup P_c$$
$$= P_b + P_o + P_s + P_c - (P_bP_o + P_bP_s + P_bP_c + P_oP_s + P_oP_c + P_sP_c)$$
$$= 0.090 - 0.0029 = \cdot 0871$$

Alternative solution

$P_f = P_b \cup P_o \cup P_s \cup P_c$ and, therefore, $\overline{P_f} = \overline{P_b \cup P_o \cup P_s \cup P_o}$

From De Morgan's Law: $= \overline{P_b \cup P_o \cup P_s \cup P_c} = \overline{P_b} \cap \overline{P_o} \cap \overline{P_s} \cap \overline{P_c}$

$$= .97 \times .99 \times .98 \times .97 = 0.91286$$

Now $Pf = 1 - \overline{P_f} = 1 - .91286 = 0.0871$

Conditional probability

When two events are referred to as being independent it means that the happening of one of these events will have no effect on the probability of happening of the other.

However there are many cases when the happening of an event can have a direct effect on the probability of the happening of a further event.

A simple illustration is the drawing of an ace from a pack of 52 playing cards on the second draw.

Let the probability of drawing an ace on the first draw be $P[A]$ and let the probability of drawing an ace on the second draw be $P[B]$

$$P[A] = 4/52 = 1/13$$

And $P[B] = 4/51$ (if there was no ace on the first draw)

but $P[B] = 3/51$ (if there had been an ace on the first draw).

In such a situation we are forced to use another symbol, $P[B|A]$, in place of $P[B]$ where $P[B|A]$ represents the value of $P[B]$ after, and knowing the result of, event A.

$P[B|A]$ is known as the conditional probability of B.

It is seen, therefore, that the formula for the probability of happening of two events, A and B, is more properly stated as:-

$$P[AB] = P[A]P[B|A]$$

and, if required, can equally well be written as:-

$$P[AB] = P[B]P[A|B] \quad (\text{ as } P[BA] = P[AB] \text{ }).$$

When we have a set of dependent events the probability that all these events will occur can be evaluated by the use of conditional probabilities.

For example, for three events A, B and C:-

$$P[ABC] = P[AB|A]P[C|AB]$$

Which is the mathematical way of saying:-

"The probability that events A, B and C will all happen is equal to the probability of event A times the probability of B, knowing the result of event A, times the probability of C knowing the result of events A and B".

A further definition of independence

Obviously if the events A, B and C are independent then $P[B|A]$ is equal to $P[B]$ and $P[C|AB]$ equals $P[C]$ so that the formula becomes:- $P[ABC] = P[A]P[B]P[C]$

This leads to a further definition of independence. If A and B are two events such that $P[B|A] = P[B]$ then A and B are statistically independent.

Example 1.8

a) Determine the probability that, during the thowing of two fair dice, either the score will be not less than 9 and/or the difference between the two individual die scores will not be less than 2.

b) Determine the conditional probability that the difference between the two die scores will not be less than 2 if the total score is not less than 9.

Solution

Let $P[A]$ = the probability that the score is not less then 9.

Let $P[B]$ = the probability that the individual scores differ by not less than 2.

The sample space of the 36 possible scores is set out below.

$\underline{6,1}$	$\underline{6,2}$	$(\underline{6,3})$	$(\underline{6,4})$	(6,5)	(6,6)
$\underline{5,1}$	$\underline{5,2}$	$\underline{5,3}$	(5.4)	(5,5)	(5,6)
$\underline{4,}1$	$\underline{4,2}$	4,3	4,4	(4,5)	$(\underline{4,6})$
$\underline{3,}1$	3,2	3,3	3,4	$\underline{3,5}$	$(\underline{3,6})$
2,1	2,2	2,3	$\underline{2,4}$	$\underline{2,5}$	$\underline{2,6}$
1,1	1,2	$\underline{1,3}$	$\underline{1,4}$	$\underline{1,5}$	$\underline{1,6}$

The elements representing event A, 6,3; 6,4;6,5;..etc. are in brackets whereas the elements representing event B are underlined.

By enumeration:- $P[A] = \frac{10}{36}$ and $P[B] = \frac{20}{36}$

a) $P[A \cup B]$, the probability that either event A or event B or both events A and B will occur can be found from enumeration and equals $\frac{26}{36}$.

b) $$P[AB] = \frac{4}{36} \quad \text{(by enumeration)}$$

$$\text{Now } P[AB] = P[A]P[B|A]$$

i.e. $$\frac{4}{36} = \frac{10}{36} P[B|A]$$

Hence $P[B|A] = \frac{2}{5}$

We can check the value obtained for $P[B|A]$ by substitution in the union probability formula:-

$$P[A \cup B] = P[A] + P[B] - P[A]P[B|A]$$

$$= \frac{10}{36} + \frac{20}{36} - \frac{10}{36} \times \frac{2}{5} = \frac{26}{36}$$

<u>The theorem of total probability</u>

Consider a set of events B_1, B_2, B_3, $\ldots B_n$ which are both mutually exclusive and also collectively exhaustive (i.e. one of the events will occur).

Then P[A], the probability of another event, A, can be expressed as:-

$$P[A] = \sum_{i=1}^{n} P[A|B_i]P[B_i]$$

Proof

The events B_i are collectively exhaustive,

i.e. $B_1, B_2, B_3, \ldots\ldots\ldots B_n = S$

$$\text{Now } P[A] = P[A]P[S] \quad (\text{as } P[S] = 1)$$

$$= P[A]P[B_1+B_2+B_3+\ldots\ldots B_n]$$

$$= P[AB_1]+P[AB_2]+P[AB_3]+\ldots P[AB_n]$$

Fig.1.5 shows the Venn diagram representing the mutually exclusive and collectively exhaustive events $B_1, B_2, B_3, \ldots B_n$. It can be seen that event A intersects these events so that the events $AB_1, AB_2, AB_3, \ldots\ldots\ldots AB_n$ are also mutually exclusive.

$$\text{Hence } P[A] = P[AB_1]+P[AB_2]+P[AB_3]+\ldots P[AB_n]$$

$$\text{Proving that } P[A] = \sum_{i=1}^{n} P[AB_i] = \sum_{i=1}^{n} P[A|B_i]P[B_i]$$

Example 1.9

A building contractor requires a roll of roofing felt.

There are three suppliers in the area and the probabilities (based on his previous experiences and the location of the suppliers), that the contractor will instruct his vanman to visit a particular supplier are:-

A - the vanman goes to supplier A : P[A] = 0.6

B - the vanman goes to supplier B : P[B] = 0.2

C - the vanman goes to supplier C : $P[C] = 0.2$

Each supplier stocks roofing felt produced by two manufacturers X and Y. Both types of roofing felt sell at the same price and both satisfy the current building regulations.

The stock situation at each of the suppliers is:-

Supplier	No. of 'X' rolls	No. of 'Y' rolls
A	10	30
B	30	20
C	30	10

The vanman will be told by his employer which supplier to visit but he will randomly select any roll of felt from the stock at the supplier.

Which roll type is the vanman most likely to return with?

Solution

Let $P[X]$ = probability that the vanman will return with roll type X.

Then, considering the stock position of each supplier:-

$$P[X|A] = 10/40 = 0.25$$

$$P[X|B] = 30/50 = 0.6$$

$$P[X|C] = 30/40 = 0.75$$

From the law of total probability:-

$$P[X] = P[X|A]P[A] + P[X|B]P[B] + P[X|C]P[C]$$

$$= 0.25 \times 0.6 + 0.6 \times 0.2 + 0.75 \times 0.2$$

$$= 0.42$$

If $P[Y]$ = probability of obtaining type Y, it can be shown in a similar manner that $P[Y] = 0.58$.

Hence it is more likely that the vanman will return with a roll of type Y.

Alternative solution - by use of a tree diagram.

The tree diagram, so named because of its appearance, can often be useful in decison making. Each branch illustrates the path that will be taken whenever a particular decision is made.

The tree diagram for this particular problem is shown in Fig.1.6 and shows all the possible combinations of events that could be involved in finishing up with an X or a Y type of roofing felt.

Tree diagrams are based on the theorem of total probability and can often be of assistance in a complex decision problem by representing it in a graphical form.

Example 1.10

If, in example 1.9, the vanman returned with an X type of roofing felt, determine the probability that he obtained it from supplier B.

Solution

The problem is simply to determine the value of $P[B|X]$

$$P[B|X] = \frac{P[XB]}{P[X]}$$

$$\text{Now } P[XB] = P[BX] = P[B]P[X|B]$$

$$\text{Hence } P[B|X] = \frac{P[B]P[X|B]}{P[X]}$$

$$= \frac{0.2 \times 0.6}{0.42}$$

$$= 0.286$$

This is an application of Bayes' theorem, which gives a relationship between prior and posterior probabilities.

BAYES' THEOREM

We have seen that:

$$P[Bi|A] = \frac{P[BiA]}{P[A]} = \frac{P[ABi]}{P[A]}$$

Now $$P[A|Bi] = \frac{P[ABi]}{P[B_i]}$$

Hence $$P[B_i|A] = \frac{P[A|Bi]P[Bi]}{P[A]}$$

and, from the theorem of total probability:-

$$P[A] = \sum_{i=1}^{n} P[A|B_i]P[B_i]$$

Therefore $$P[B_i|A] = \frac{P[A|Bi]P[Bi]}{\sum_{i=1}^{n} P[A|B_i]P[B_i]}$$

The above expression is known as Bayes' theorem (or rule) and can be expressed in words as:-

If A is an event that could be caused by any one of n different events, B_i, all of which are both mutually exclusive and collectively exhaustive ($\sum_{i=1}^{n} B_i = S$), then Baye's theorem gives the relationship between the probability of event B_i happening (given the result of the happening of event A) to the probability of A happening (given the result of the happening of event B_i).

$P[B_i]$ = prior probability of event B_i (with no knowledge of A)

$P[B_i|A]$ = posterior probability of event B_i (after noting A)

$P[A|B_i]$ = likelihood of event A (after noting B_i)

Example 1.11

Solve example 1.10 using Bayes' theorem.

Solution

Let events A, B and C of example 1.9 be B_1, B_2 and B_3, so that $P[B|X]$ of example 1.10 is expressed as $P[B_2|X]$.

By Bayes´ theorem :-

$$P[B2|X] = \frac{P[X|B2]P[B2]}{\sum_{i=1}^{3} P[X|B_i]P[B_i]} = \frac{0.6\text{x}0.2}{0.25\text{x}0.6 + 0.6\text{x}0.2 + 0.75\text{x}0.2}$$

$$= 0.286$$

Note

Bayes´ theorem is extremely useful as a technique to continually process information.

In any probability analysis the designer must make assumptions in order to have a set of prior probabilities.

Bayes´ theorem can be used to continually adjust these assumed probability values to conform with new information as it becomes available and is then usually written in the form:-

$$P[\text{state}|\text{sample}] = \frac{P[\text{sample}|\text{state}]P[\text{state}]}{\sum_{\text{all states}} P[\text{sample}|\text{state}]P[\text{state}]}$$

Example 1.12

The overall strength of an existing embankment is to be assessed in order to determine the practicality of running a roadway across it.

The soil in the embankment is largely cohesive and, after studying records of its past performance and having visited the site to check its appearance and general state of repair the soils engineer has concluded that the soil is mainly of a firm consistency.

His definitions of consistency were based on the unconfined compressive strength of the soil, c_u, as follows:-

Soft consistency .... $c_u < 24\text{kN/m}^2$

Firm .... $c_u = 24$ to 48kN/m^2

Stiff .... $c_u > 48\text{kN/m}^2$

If we let B_i = the average consistency of the soil in the embankment then B_i can be either soft, firm or stiff.

The engineer assiqned prior propabilities for B_i as :-

Average consistency (B_i)	Prior probability $P[B_i]$
Soft	0.3
Firm	0.5
Stiff	0.2

If the state of the soil in the embankment is known then it is a simple matter to predict a value for the unconfined strength, A, that one would obtain from a test on a sample. However one test can never be conclusive and the soils engineer, with the help of control tests and his experience with similar problems, evolved a set of conditional probabilities, $P[A|B_i]$, that are set our below.

A - the c_u value obtained in a test on a sample (kN/m^2).	B_i - The average state, i.e. consistency of the soil. Soft	Firm	Stiff
<24 (indicates soft - A_s)	.7	.3	0
24 to <48 (indicates firm - A_f)	.3	.6	.2
>48 (indicates stiff - A_{st})	0	.1	.8

He then slightly adjusted the $P[A|B_i]$ values to use 0.01 instead of 0 and hence avoid any multiplications by zero.

This is necessary in order to avoid the elimination of a probability that may increase as more data becomes available.

Sample A_i	State $P[A_i \mid B_i]$		
	Soft	Firm	Stiff
A_s	.7	.3	.01
A_f	.29	.6	.19
A_{st}	.01	.1	.8

If the reader is in some doubt to what exactly $P[A_i|B_i]$ means perhaps the following explanation may be of assistance.

If the state is soft, i.e. B_i = soft, then the probability of a test on a sample indicating this situation, (i.e. recording an unconfined compressive strength of $<24kN/m^2$) = 0.7.

In symbols : $P[A_s|B_s] = 0.7$

Similarly $P[A_s|B_f] = 0.3$

and $P[A_s|B_{st}] = 0.01$

where the suffices s, f and st stand for soft, firm and stiff respectively.

Obviously for a particular state a test result must reflect some consistency, either soft, firm or stiff, which explains why the $P[A|B_i]$ values in the table summate vertically to 1.0.

Assume that, in this case, four samples were collected from different locations in the embankment and that by the unconfined compression test two of the samples indicated a stiff consistency,

A_{st}, one a firm consistency, A_f, and one a soft consistency, A_s.

Determine the main consistency state of the embankment.

Solution

The prior probabilities, assumed by the engineer, for the state of the embankment are:-

$$P[B_s] = 0.3; \quad P[B_f] = 0.5; \quad P[B_{st}] = 0.2$$

Consider test 1 - result A_{st}

Having a test result, A_{st}, we can work out the posterior probabilities as to the state of the embankment:-

$P[B_s|A_{st}]$ - the probability that the consistency is soft (knowing that the test result indictes that it is stiff).

$P[B_f|A_{st}]$ - the probability that the consistency is firm.

$P[B_{st}|A_{st}]$ - the probability that the consistency is stiff.

From Bayes´ theorm:-

$$P[B_s|A_{st}] = \frac{P[Ast|Bs]P[Bs]}{P[A_{st}|B_s]P[B_s] + P[A_{st}|B_f]P[B_f] + P[A_{st}|B_{st}]P[B_{st}]}$$

$$= \frac{.01x0.3}{.01x0.3 + 0.1x0.5 + .8x0.2} = \frac{.003}{.213} = .0141$$

$$P[B_f|A_{st}] = \frac{0.1x0.5}{.213} = .2347; \quad P[B_{st}|A_{st}] = \frac{.8x0.2}{.213} = .7512$$

Hence, prior to considering the results of the second test, we have upgraded the $P[B_i]$ values to:-

$$P[B_s] = .0141; \quad P[B_f] = .2347; \quad P[B_{st}] = .7512$$

The process continues in an identical manner:-

Test 2 - result A_{st}

$$P[B_s|A_{st}] = \frac{.01x0.0141}{01x0.0141 + 0.1x0.2347 + .8x0.7512} = \frac{.0001}{.6246} = .0002$$

$$P[B_f|A_{st}] = \frac{0.1x0.2347}{.6246} = .0376; \quad P[B_{st}|A_{st}] = \frac{.8x0.7512}{.7512} = .9622$$

Test 3 - result A_f

$$P[B_s|A_f] = \frac{.29x.0002}{29x0.0002 + .6x0.0376 + .19x0.9622} = \frac{.00006}{.2054} = .0002$$

$$P[B_f|A_f] = \frac{.6x0.0376}{.2054} = .1097; \quad P[Bst|Af] = \frac{.19x0.9622}{.2054} = .8900$$

Test 4 - result A_s

$$P[B_s|A_s] = \frac{.7x0.0002}{.7x0.0002 + .3x0.1097 + .01x0.8900} = \frac{.00014}{.0420}$$

$$= .0033$$

$$P[B_f|A_s] = \frac{.3x0.1097}{.0420} = .7845; \quad P[Bst|As] = \frac{.01x0.8900}{.0420} = .2122$$

Probabilities of state of embankment are:-

Soft consistency P = 0.3%

Firm consistency P = 78.5%

Stiff consistency P = 21.2%

The reader might like to check that the final probabilities are unaffected by the order in which the test results are considered.

Alternative solution

The foregoing procedure was listed in full as a demonstration but it is not necessary to consider each test result separately.

As the test results are independent the values of the conditional probabilities of the results, $P[A|Bi]$, are equal to the product of the four conditional probabilities:-

$$P[A|B_s] = P[A_{st}|B_s]P[A_{st}|B_s]P[A_f|B_s]P[A_s|B_s]$$
$$= 0.01x0.01x0.29x0.7 = 0.0000203$$

$$P[A|B_f] = P[Ast|Bf]P[Ast|Bf]P[Af|Bf]P[As|Bf]$$
$$= 0.1x0.1x0.6x0.3 = 0.0018$$

$$P[A|B_{st}] = P[A_{st}|B_{st}]P[A_{st}|B_{st}]P[A_f|B_{st}]P[A_s|B_{st}]$$

$$= 0.8x0.8x0.19x0.01 = 0.001216$$

Now, from Bayes´ theorm:-

$$P[B_s|A] = \frac{P[A|Bs]P[Bs]}{P[A|B_s]P[B_s] + P[A|B_f]P[B_f] + P[A|B_{st}]P[B_{st}]}$$

The prior probabilities for B_s, B_f and B_{st} were 0.3, 0.5 and 0.2

Hence $P[B_s|A] = \frac{.0000203x0.3}{.0000203x0.3 + .0018x0.5 + .001216x0.2} = .0053$

and $P[B_s|A] = \frac{.0018x0.5}{.001149} = .7831$; $P[Bst|A] = \frac{.00122x0.2}{.001149} = .2116$

Giving $P[B_s] = 0.5\%$; $P[B_f] = 78.3\%$ and $P[B_{st}] = 21.2\%$

Note

Example 1.10 indicates the possible dangers of adopting a deterministic approach in solving a civil engineering problem.

The soils engineer is experienced, knows the site, the testing techniques and the personnel involved.

His judgement has been included in a probability analysis which indicates strongly that the soil in the embankment is almost entirely of a firm consistency.

If the test results alone are considered then there is every likliehood that the conclusion will be that the embankment consists of predominantly stiff soil.

Example 1.13

In the clay lens problem of example 1.6 the engineer assesses that there is a 50% chance of a clay lens being within the site area.

He also estimates that, if there is a lens, the chance of encountering it in a borehole is 0.1, (See method in Example 1.6).

Determine the change in this probability if, after 5 boreholes, no clay has been encountered.

Solution

There are two states :- 1 - Lenses present; 2 - No lenses

There are two test results (samples):- 1 - no find; 2 - a find

State - lenses present

Probability of a find in one borehole, P[find] = 0.1

Hence P[no find] = 1 - P[find] = 1 - 0.1 = 0.9

Hence probability of no find in 5 boreholes = 0.9^5

State - no lenses present

Obviously P[find] = 0 and P[no find] = 1

Using Baye´s theorem:-

P[lenses|no find] =

$$\frac{P[\text{no find}|\text{lenses}] \times P[\text{lenses}]}{P[\text{no find}|\text{lenses}] \times P[\text{lenses}] + P[\text{no find}|\text{nolenses}] \times P[\text{no lenses}]}$$

$$= \frac{0.9^5 \times 0.5}{0.9^5 \times 0.5 + 1 \times 0.5} = .371$$

The probability of there being clay lenses within the site area has reduced from 50 to 37%.

Note

The same result is obtained if the calculation is carried out for the five separate events of not finding clay in each borehole, in a similar manner to the way example 1.12 was first solved. However, apart from the extra work involved, it is necessary to work to several places of decimals to avoid significant rounding off errors.

CHAPTER TWO - THE SECOND MOMENT METHOD OF RELIABILITY ANALYSIS

The probability of failure of a structure

The term "failure" is used here in its most general sense and implies the failure of the structure to satisfy some particular limit state criterion, which may or may not be actual structural failure.

The frequenistic approach cannot be applied to the estimation of the probability of failure of a civil engineering structure where the design and construction is a once only operation.

Even for similar, or prefabricated structures, where aspects of the design work may be repeated, each structure will be built on a different site leading to the possibility of different soil and geological conditions.

An evaluation of the probability of failure of a structure must therefore be undertaken by the application of statistics and probability theory.

Methods of reliability analysis

There are three main methods by which a structure may be designed to achieve a certain probability of failure value and these are described in Report 63 of C.I.R.I.A., (1976):-

Level I - A design method involving characteristic values and partial factors of safety.

Level II - A reliability analysis which uses safety checks at a selected point (or points) on the failure boundary, defined by the appropriate limit state function, Z.

Level III - an extrememely comprehensive probabilistic analysis in which safety checks based on 'exact' probabilistic analyses, using a full distributional approach, are carried out for the whole structural system.

The level I approach is more of a hope for the future than a method that exists at the moment.

If ever evolved the method will not require the actual evaluation of P_f. A particular limit state will be considered safe if the appropriate partial safety factors are not exceeded. A list of these factors will be given in the design codes.

The method would generally involve structural design calculations very similar to those produced at present. The final design proposals would be probabilistically based even if the design engineer did not have a comprehensive knowledge of probability theory.

The problem is that values for these partial safety factors, for a suitable range of structural elements, will first have to be obtained from reliablity analsyes carried out by either a Level II or a Level III approach.

The level III method is really a form of pure mathematics and will probably only be used for the reliability analysis of special structures, in which the reliability level is of critical importance or where it is particularly important to optimise the design.

Level II methods involving fairly straightforward mathematics can now be evolved and there is general agreement that Level II methods have the potential of either being used to evaluate suitable partial factors for use in Level I analyses or being used directly as design methods in their own right.

This thesis will deal mainly with the Level II method.

RANDOM VARIABLES

In order to use probability theory it is necessary to express the engineering uncertainties in terms of numerical values which can then be considered to vary in an uncertain, or random, way.

For example if, on a particular site, the unit weight of the soil varies from 18 to 20kN/m^3 it is not possible to fix on an actual numerical value for this parameter.

The procedure adopted is to designate variable values by capital letters A, B, C,...etc. signifying that the particular value of the parameter represented by the letter is not constant but varies randomly over a range of possible values, in our case from 18 to 20kN/m^3.

Lower case letters a, b, c, ... are generally used to denote the various values that the random variables A, B, C, ... can have.

A particular form of the random process is the stochastic process in which the rate of occurrence of different values follows some sort of statistical pattern. In such a case the variable concerned, although still correctly defined as random, is sometimes referred to as a stochastic variable.

If the values of the variable can only be from a finite distribution of values, eg. the integer scores possible with two dice, the variable is referred to as a discrete random variable.

If the values of the variable are continuous, eg. 6.00001 is considered different to 6.00002, then the variable is referred to as a continuous random variable.

Engineering situations are somewhat removed from dice throwing and most engineering probability distributions are continuous, although the process of rounding off to so many decimal places and the limitations of the measuring apparatus often leads to a set of measured values that appear to be from a discrete distribution.

Consider the resistance or strength of a structure, R and the applied loading, S, to which it will be subjected.

The values of both R and S are not fixed but will assume any value within a range of values. The extent of these ranges will vary with the degree of probability decided as acceptable for the design problem, (usually 95%).

R and S are therefore random variables with definitive, although possibly unknown, probability density functions (p.d.f.s).

Figs.2.1A and B show assumed p.d.f.s for R and S and illustrate that failure will occur when $R < S$.

If Fig.2.1B is subtracted from Fig.2.1A then the probability curve of $Z = R - S$, (strength minus load), is obtained, Fig.2.1C.

The probability of failure, $P_f = P[(R-S)=0] = P[Z = 0]$

Reliability Index

Generally there is not sufficient information regarding the tails of the Z distribution and the criterion $P_f = P[Z = 0]$ is therefore replaced with one that involves the mean value and standard deviation of Z.

In Fig.2.1C the distance from the mean of Z, m_z, to the failure boundary, i.e. the point at which Z = 0, can be expressed in terms of σ_z, the standard deviation of Z, and equals $\beta\sigma_z$. β is known as the reliability index and is a measure of the safety of the system.

Obviously $m_z - \beta\sigma_z = 0$ i.e. $\beta = m_z/\sigma_z$

Now $m_z = m_R - m_S$ Hence $\beta = \dfrac{m_R - m_S}{\sigma_z}$

The factor of safety, F, is equal to m_R/m_S.

The expression for F is purely deterministic whereas the expression for β includes not only m_R and m_S but also σ_z, a measure of the uncertainty of both R and S. It can therefore be seen that is a more meaningful measure of reliability than F.

<u>Basic variable space</u>

In most practical problems R and S will rarely be single variables and will be vectors made up from the set of relevant basic variables.

Basic variables are the fundamental parameters involved in the design. Examples are the ultimate strengths of the materials to be used, the intensity and type of loadings, depth of reinforcement,etc.

If n basic variables make up a particular random variable X such that:-

$$X = (X_1, X_2, X_3, \ldots\ldots\ldots\ldots X_n)$$

then the basic variable space is the n dimensional space that will represent all possible values of X.

This means that $x = (x_1, x_2, x_3, \ldots\ldots\ldots x_n)$ is a single point of coordinates $x_1, x_2, x_3, \ldots\ldots x_n$ and represents the situation when the basic variables $X_1, X_2, X_3, \ldots . X_n$ have values x_1 to x_n.

Z is a function of all the relevant basic variables so we can say that, generally:-

$$Z = g(X_1, X_2, X_3, \ldots\ldots\ldots X_n)$$

and that $P_f = P[Z = 0] = P[g(X_1, X_2, X_3, \ldots\ldots\ldots X_n) = 0]$

Example 2.1

A granular soil will be subjected to a shear stress, τ.

The normal stress on the shear plane, σ, will have a mean value of 100kN/m^2 and a standard deviation of 20 kN/m^2.

The angle of friction of the soil has a mean value of 35° and a standard deviation of 5°.

Plot the failure boundary in the basic variable space and determine the reliablity index of the system if τ has a fixed value of 50kN/m^2.

Solution

Coulomb's Law of soil shear strength states that:-

$$\tau = \sigma \tan\phi \quad \text{for a granular soil}$$

$$\Rightarrow Z = \sigma\tan\phi - \tau = \sigma\tan\phi - 50$$

Z can be represented as $Z = g(\sigma, \phi, \tau)$ or $Z = g(X_1, X_2, C)$

where $X_1 = \sigma$, $X_2 = \phi$ and C = a constant (= 50).

As there are only two random variables n = 2. The failure boundary will show up as a line on a two dimensional plot. This line can be obtained from the equation $\tan\phi = 50/\sigma$ which is purely deterministic and has nothing to do with probability theory.

However if the scales for σ and ϕ are so chosen that the length representing one standard deviation of σ (20kN/m^2) is equal to the length that represents one standard deviation of ϕ (5°) then the minimum distance from the mean point to the failure boundary will be equal to β, the reliability index.

By selecting suitable values for σ a range of corresponding $\tan\phi$ values can be obtained which leads to the values of ϕ tabulated below.

σ(kN/m^2)	40	60	80	100	120	140	160
$\tan\phi$	1.25	.833	.625	.5	.417	.357	.313
ϕ(degs.)	51.3	39.8	32.0	26.6	22.6	19.6	17.4

The failure boundary is shown in Fig.2.2A together with the mean point (100,35°). The central part of the diagram is shown enlarged in Fig.2.2B so that an accurate determination of the minimum distance from the mean point to the failure boundary can be obtained. Due to the scales chosen this distance is in terms of the standard deviation of Z and is found to be equal to 1.16. Hence the reliability index of the system is 1.16.

Note that the smaller the minimum distance the nearer the mean point is to the failure boundary and the greater the risk of failure.

The second moment method of reliability analysis

When a number of variables is involved the failure boundary is a surface, not a line, and the plotting technique just described cannot be used. The approach, therefore, has to be mathematical.

As discussed the most practical method is the Level II approach which, as it deals with means and variances, is classified as a second moment approach. (The variance of a random variable is its second central moment).

Second moment methods of reliability analysis orginate from work by Mayer, (1926), but were not seriously considered for a further forty years when the works of Cornell, (1969), Ravindra et al.,(1969) and Rosenblueth and Esteva, (1972) were published.

The proposed techniques suggested that a simple method of safety checking, involving some statistical measure of the uncertaintites involved but without employing complex integrations using full probability distributions could be evolved.

The mathematics are considerably simplified if the failure surface is linear. Such a situation rarely occurs (see for example Fig.2.2) but it is possible to obtain a local approximation to linearity by means of a Taylor's expansion in which 2nd order terms and above are ignored. Because of this the method is usually referred to as a first- order second moment method.

<u>The mean value first-order second moment method</u>

This method, reported on by Cornell, (1969) and Rosenblueth and Estava, (1972), consists of expanding the limit state function at the mean point, in order to create the local approximation to a linear failure surface.

It has been established that the failure, i.e. the limit state, of a structure can be expressed as a function of the relevant basic variables:-

$$Z = g(X) = g(X_1, X_2, X_3, \ldots\ldots.X_n)$$

Consider first the case of $Z = g(X)$ where X is a single variable, (Fig.2.3). Then, using Taylor's expansion:-

$$Z \doteqdot g(a) + (x - a)g'(a) + \frac{(x - a)^2}{2}g''(a) + \ldots\ldots\ldots$$

where a = the value of X at which the approximation is taken.

$$g'(a) = \left.\frac{dg(X)}{dX}\right|a = \text{the first derivative of } g(X) \text{ evaluated for } X = a.$$

Removing second order terms and above results in a first order approximation, consisting of two terms:-

$$Z \doteq g(a) + (x - a)g'(a)$$

If Z is a function of several variables the equivalent expression is:-

$$Z \doteq g(a_1, a_2, a_3, \ldots.a_n) + \sum_{i=1}^{n} (x_i - a_i)g_i(a)$$

If the expansion is to take place at the mean point then all that is necessary is to change the "a" in the above expression to "m".

$$Z \doteq g(m_1, m_2, m_3, \ldots\ldots..m_n) + \sum_{i=1}^{n} (x_i - m_i)g_i'(m_i)$$

Example 2.2

By expanding about the mean values determine an expression for the lineal approximation for the failure boundary of example 2.1

Solution

The expression for Z is: $Z = \sigma\tan\phi - \tau = 0$

$$\text{i.e. } Z = g(X_1, X_2, C)$$

where $X_1 = \sigma$ (Mean = 100kN/m^2; Standard deviation = 20kN/m^2)

$X_2 = \phi$ (Mean = 35°; Standard deviation = 5°)

$C = \tau$ (a constant = 50kN/m^2)

For X_1 $\quad g'(X_1) = \left.\frac{\partial Z}{\partial\sigma}\right|_{m_\sigma;m_\phi} = \tan\phi = 0.7002$

For X_2 $\quad g'(X_2) = \left.\frac{\partial Z}{\partial\phi}\right|_{m_\phi;m_\sigma} = \sigma\sec^2\phi = 149.03$

$$\text{Hence } Z = m_1\tan m_2 - C + (X_1 - m_1)g'(X_1) + (X_2 - m_2)g'(X_2)$$

$$= 20.02 + 0.7002(X_1 - 100) + 149.03(X_2 - 35^\circ)\frac{\pi}{180}$$

$$\text{i.e. } X_1 + 3.715X_2 - 201.424 = 0$$

Selecting suitable values for X_1, (σ), leads to the values of ϕ (in degrees) tabulated below:-

σ (kN/m^2)	40	60	80	100	120	140
ϕ	43.5	38.1	32.7	27.3	21.9	16.5

The lineal approximation of the failure surface is shown plotted as a dashed line in Fig.2.2. It is seen that the approximation is not all that good in that, within the region of the mean point, the line tends to lie within the safe zone.

By scaling the minimum distance from the mean point to the approximated failure boundary the value of β is found to be 1.04.

Obviously in a simple two dimensional problem where the failure boundary can be plotted directly, as in example 2.1, there is no need to use any form of approximation. However reliability analyses invariably involve several basic variables and some form of mathematical analysis, such as the second moment approach, is necessary.

The purpose of example 2.2 is to illustrate that although the method of mean second moments is very simple the fact that linearisation takes place at the mean point tends to place this approximation within the safe zone which can lead to an unrealistically low value for β.

An even more important problem is that of invariance, which is illustrated by the following example.

Example 2.3

A short column has a diameter X_1 and is loaded with an axial compressive load X_2. The ultimate compressive stress of the column is X_3.

The variables have the following mean and s.d. values:-

	mean	s.d.
X_1	3.5	0.4
X_2	10.0	1.0
X_3	2.5	0.5

Determine the reliability index of the system.

Solution

$Z = R - S$ so the limit state function can be written as:-

$$Z = g(X) = X3\ \frac{\pi}{4}\ X_1^2 - X_2 = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

or as $$Z = g(X) = \frac{\pi}{4}\ X_1^2 - \frac{X_2}{X_3} = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

or as $$Z = g(X) = \frac{X_3}{X_2}\ \frac{\pi}{4}\ X_1^2 - 1 = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Now it can be shown that, Benjamin & Cornell, (1970):-

If $Z = g(X)$ where $X = (X_1, X_2, X_3, \ldots\ldots X_n)$

$$\text{then} \quad \sigma_Z = \sqrt{\sum_{i=1}^{n} (g_i{}' \sigma_X)^2}$$

where $g_i{}' = \left.\frac{\partial Z}{\partial X}\right| m_i$ (the first derivative of Z with $X = m_i$.

Hence, for equation (1):-

$$m_Z = \frac{2.5\pi 3.5^2}{4} - 10 = 14.05$$

$$\sigma_Z = \sqrt{\left(\frac{m_3 \pi m_1}{2}\sigma_1\right)^2 + \left(-m_2\sigma_2\right)^2 + \left(\frac{\pi m_1{}^2}{4}\sigma_3\right)^2}$$

$$= \sqrt{\left(\frac{2.5\pi 3.5}{2}x0.4\right)^2 + \left(-10x1\right)^2 + \left(\frac{\pi 3.5^2}{4}x0.5\right)^2} = 12.38$$

$$= \frac{14.05}{12.38} = 1.14$$

The reader might like to check that the β values for equations (2) and (3) work out as 1.27 and 2.51 respectively.

The advanced first order second moment method

The above example has illustrated the main drawback of the mean value second moment approach. This is that the position of the boundary approximation can vary with the Z formulation. Hence it is possible to obtain two different values for β, a fact that makes the method virtually unacceptable.This sensitivity of the reliability index was pointed out by Ditlevsen, (1973) and an invariant second moment index was proposed by Hasofer and Lind ,(1974).

Hasofer and Lind showed that an invariant reliability index is obtained if the point chosen for the lineal approximation is actually on the failure boundary. This point of maximum probability of failure, generally called the design point, is given the symbol x^* and lies somewhere along the boundary.

Hasofer and Lind´s work was extended by Rackwitz, (1976) and has led to the advanced first order second moment method, the principles of which are described below.

If we assume that a particular limit state function $Z = g(X)$ consists of the single variable, X, the failure boundary can then be plotted, (Fig.2.4).

From the plot it is seen that at x^* the value of $Z = g(x^*) = 0$.

With more than one variable $g(x^*) = g(x_1^*, x_2^*, x_3^*, \ldots\ldots\ldots x_n^*)$ where $x_1^*, x_2^*, x_3^* \ldots\ldots$ are the design values of $X_1, X_2, X_3, \ldots$etc.

An expression for the linear approximation of the failure boundary at $a = a_1, a_2, a_3, \ldots\ldots a_n$ was established earlier:-

$$Z \doteq g(a_1, a_2, a_3, \ldots\ldots\ldots a_n) + \sum_{i=1}^{n} (x_i - a_i) g_i(a)$$

and, by substituting "x^*" for "a" in the above expression the failure boundary approximation at the design point is found to be:-

$$Z \doteqdot \sum_{i=1}^{n} (x_i - x_i^*)\, g_i(x^*) \qquad (\text{as } g(x_1^*, x_2^*, x_3^*, \ldots\ldots x_n^*) = 0)$$

$$\Rightarrow \quad m_Z \doteqdot g(x_1^*, x_2^*, x_3^*, \ldots\ldots\ldots x_n^*) + \sum_{i=1}^{n} (m_i - x_i^*) g_i'(x^*)$$

$$= \sum_{i=1}^{n} (m_i - x_i^*) g_i'(x^*)$$

$$\text{and } \sigma_Z \doteqdot \sqrt{\sum_{i=1}^{n} [g_i'(x^*)\sigma_i]^2}$$

where $g_i'(x^*)$ = the first derivative of $g(X)$ evaluated at the point $x^* = (x_1^*, x_2^*, x_3^*, \ldots\ldots x_n^*)$

The sensitivity factor

A measure of the contribution of any variable, X_i, to the value of σ_Z is its sensitivity factor, α_i, which is simply the ratio:-

$$\alpha_i = \frac{g_i'(x^*)\sigma_i}{\sigma_Z}$$

$$\text{Now} \quad \sigma_Z^2 \doteqdot \sum_{i=1}^{n} [g_i'(x^*)\sigma_i]^2$$

$$= \sum_{i=1}^{n} (\alpha_i \sigma_Z) g_i'(x^*)\sigma_i$$

$$= \sigma_Z \sum_{i=1}^{n} \alpha_i g_i'(x^*)\sigma_i$$

$$\text{Hence } \sigma_Z \doteqdot \sum_{i=1}^{n} \alpha_i g_i'(x^*)\sigma_i$$

$$\text{Now} \quad \beta = \frac{m_Z}{\sigma_Z} = \frac{\sum_{i=1}^{n} (m_i - x_i^*) g_i'(x^*)}{\sum_{i=1}^{n} \alpha_i g_i'(x^*)\sigma_i}$$

Therefore $$\sum_{i=1}^{n} g_i'(x^*)\,[(m_i - x_i^*) - \alpha_i \beta \sigma_i] = 0$$

The value of x_i^* that satisfies this equation is given by:-

$$x_i^* = m_i - \alpha_i \beta \sigma_i \text{ for all values of } i.$$

By determining all values of x_i^* the design point x^* can be obtained. The solution technique givem in C.I.R.I.A.'s Report 63, (1976) is as follows:-

1. Guess a value for β
2. Set $x_i^* = m_i$ for all i values
3. Compute $\partial g/\partial x_i$ for all i, at $x = x^*$
4. Compute α_i for all i
5. Compute new x_i^* values
6. Repeat steps 3 to 5 until stable values of x_i^* are achieved
7. Evaluate $Z = g(x_1^*, x_2^*, x_3^*, \ldots\ldots\ldots\ldots x_n^*)$
8. Modify β and repeat steps 3 to 7 to achieve Z = 0

Example 2.4

Show that a value of 1.156 for β determines the design point of example 2.1.

Solution

$$Z = \sigma \tan\phi - \tau = 0$$

$$= g(X1, X2, C) \quad \text{where } X1 = \sigma;\ X2 = \phi;\ C = 50\text{kN/m}^2$$

Setting x_i^* values equal to m_i gives $x_1^* = 100\text{kN/m}^2$ and $x_2^* = 35^o$

$$g_1'(x^*) = \frac{\partial Z}{\partial X_1} = \tan\phi|x_1^*;x_2^* = .7002$$

$$g_2'(x^*) = \frac{\partial Z}{\partial X_2} = \sec^2\phi|x_1^*;x_2^* = 149.03$$

With these values $\sigma_Z = \sqrt{\left(\frac{\partial g}{\partial X_1}\sigma_{X_1}\right)^2 + \left(\frac{\partial g}{\partial X_2}\sigma_{X_2}\right)^2}$

$$= \sqrt{(.7007 \times 20)^2 + (149.03 \times .0873)^2}$$

$$= 19.1149$$

$$\alpha_1 = \frac{g_1'(x^*)\sigma_{X1}}{\sigma_Z} = \frac{.7007 \times 20}{19.1149} = 0.7326$$

$$\alpha_2 = \frac{g_2'(x^*)\sigma_{X2}}{\sigma_Z} = \frac{149.03 \times .0873}{19.1149} = 0.6806$$

$$\Rightarrow x_1^* = m_1 - \alpha_1 \beta \sigma_{X1} = 100 - .7362 \times 1.156 \times 20 = 83.06$$

$$x_2^* = m_2 - \alpha_2 \beta \sigma_{X2} = .6109 - .6806 \times 1.156 \times .0873 = .5422 \text{ rads.}$$

$$\Rightarrow x_2^* = 31.06^\circ$$

The full iteration is set out below:-

Iteration	Variables	$g_i'(x^*)$	α_i	x_i^*
START	$X_1 = \sigma$			$100 kN/m^2$
	$X_2 = \phi$			35°
1	X_1	.7002	.7328	83.06
	X_2	149.03	.6806	31.06
2	X_1	.6024	.7733	82.12
	X_2	113.20	.6340	31.34
3	X_1	.6089	.7783	82.01
	X_2	112.56	.6279	31.37
4	X_1	.6097	.7790	81.99
	X_2	112.49	.6271	31.38
5	X_1	.6098	.7790	82.00
	X_2	112.48	.6270	31.38

Using the final derived values for X_1 and X_2 the closing error for

Z can be obtained and equals 0.0137, a value that most engineers would accept as equivalent to zero.

Therefore:- Design point = ($82 kN/m^2$, 31.38^o) and $\beta = 1.156$

Note

It is unfortunate for geotechnical engineers that the symbol used in statistics for the standard deviation is the same as that used for normal stress, σ. It seems best to leave the two terms with the same symbol rather than to change one. In most cases there should be little risk of ambiguity.

Example 2.5

The invariance of β, when obtained by the advanced order method, will now be illustrated by re-solving example 2.3.

The three forms of the limit state function obtained were:-

$$Z = g(X) = X3 \frac{\pi}{4} X_1^2 - X_2 = 0 \quad \text{.....................(1)}$$

$$Z = g(X) = \frac{\pi}{4} X_1^2 - \frac{X_2}{X_3} = 0 \quad \text{.......................(2)}$$

$$Z = g(X) = \frac{X_3}{X_2} \frac{\pi}{4} X_1^2 - 1 = 0 \quad \text{....................(3)}$$

Using each of these equations in turn an iterative procedure was adopted and showed that, in each case, β equals 2.41.

The complete set of iterations is tabulated below.

Var.		Equation (1) $g_i'(x^*)$	α_i	x_i^*	Equation (2) $g_i'(x^*)$	α_i	x_i^*	Equation (3) $g_i'(x^*)$	α_i	x_i^*
0	X_1			3.5			3.5			3.5
	X_2			10.0			10.0			10.0
	X_3			2.5			2.5			2.5
1	X_1	13.74	0.75	2.78	5.50	0.93	2.61	1.37	0.71	2.81
	X_2	-1.00	-0.14	10.33	-0.40	-0.17	10.41	-0.24	-0.31	10.75
	X_3	9.62	0.65	1.71	1.60	0.34	2.09	0.96	0.63	1.75
2	X_1	7.49	0.68	2.84	4.10	0.79	2.74	0.72	0.69	2.84
	X_2	-1.00	-0.23	10.55	-0.48	-0.23	10.55	-0.09	-0.22	10.54
	X_3	6.08	0.69	1.66	2.37	0.57	1.81	0.58	0.69	1.67
3	X_1	7.43	0.67	2.86	4.30	0.71	2.81	0.71	0.67	2.86
	X_2	-1.00	-0.22	10.54	-0.55	-0.23	10.55	-0.09	-0.22	10.54
	X_3	6.34	0.71	1.64	3.21	0.66	1.70	0.60	0.71	1.64
4	X_1	7.38	0.66	2.86	4.42	0.68	2.85	0.70	0.66	2.86
	X_2	-1.00	-0.22	10.54	-0.59	-0.23	10.54	-0.09	-0.22	10.54
	X_3	6.41	0.72	1.64	3.65	0.70	1.66	0.61	0.72	1.64
5	X_1	7.38	0.66	2.86	4.47	0.66	2.86	0.70	0.66	2.86
	X_2	-1.00	-0.22	10.54	-0.60	-0.22	10.54	-0.09	-0.22	10.54
	X_3	6.41	0.72	1.64	3.84	0.71	1.64	0.61	0.72	1.64

For all three iterations $|Z|$ is less than 0.00.

Reduced variables

It is generally more convenient to work in terms of "reduced" or "standardised" variables.

If x_1 is the particular value of a variable with a mean of m_1 and a standard deviation of σ_1, then the corresponding reduced variable, y_1, is given by the expression:-

$$y_1 = \frac{x_1 - m_1}{\sigma_1}$$

A reduced variable has the properties that its mean value is 0 and its standard deviation is 1.0 which means that the origin of the axes that represent this reduced space is also the mean point of the reduced variables.

The failure surface can now be expressed as $Z = g(y)$

where $g(y) = g(y_1, y_2, y_3, \ldots\ldots\ldots y_n)$

Taylor's first degree approximation to Z at the point x^* has already been established:-

$$Z \doteqdot g(x_1^*, x_2^*, x_3^*, \ldots x_n^*) + \sum_{i=1}^{n} (x_i - x_i^*) g'_i(x^*)$$

The linear approximation at the point $y^* = (y_1^*, y_2^*, y_3^*, \ldots y_n^*)$ is therefore:-

$$Z \doteqdot g(y_1^*, y_2^*, y_3^*, \ldots y_n^*) + \sum_{i=1}^{n} (y_i - y_i^*) g'_i(y^*)$$

simplifying to $Z \doteqdot \sum_{i=1}^{n} (y_i - y_i^*) g'_i(x^*)$

The mean of Z is therefore:-

$$m_Z \doteqdot \sum_{i=1}^{n} (m_i - y_i^*) g_i'(y^*) \quad = - \sum_{i=1}^{n} y_i^*(y^*)$$

(as the mean of a standardised variable = 0)

and $$\sigma_Z \doteqdot \sum_{i=1}^{n} \alpha_i g_i'(y^*) \sigma_i$$

$$= \sum_{i=1}^{n} \alpha_i g_i'(y^*)$$

(as the variance of a standardised variable = 1.0)

Now $$\beta = \frac{m_Z}{\sigma_Z} = \frac{- \sum_{i=1}^{n} y_i g_i'(x^*)}{\sum_{i=1}^{n} \alpha_i g_i'(y^*)}$$

Therefore $$\sum_{i=1}^{n} g_i'(y^*) \, [-(y_i^* - \alpha_i \beta] ' = 0$$

The solution, in terms of the standardised variables is therefore:-

$$y^* = -\alpha_i \beta \qquad \text{for all } i$$

From the above equation it is seen that the distance from the origin to y^* is a measure of the reliability index. It can be obtained from the expression:-

$$\beta = \sqrt{\sum_{i=1}^{n} y_i^2}$$

<u>Iterative procedure for determining β</u>

An algorithm proposed by Fiessler, (1980) can be used with reduced variables and gives the value of β after only one set of iterations.

A suitable procedure is as follows:-

1. Determine an expression for g(X).
2. Evolve an expression for g(y).
3. Determine expressions for all first derivatives of g(y), g_i'.
4. Set $y_i = 0$ and $\beta = 0$.
5. Evaluate all g_i' values.
6. Evaluate g(y).
7. Evaluate standard deviation of Z from $\sigma_z = \sqrt{\sum (g_i')^2}$
8. Evaluate new values for y from $y = \dfrac{-g_i'}{\sigma_z}\left[\beta + \dfrac{g(y)}{\sigma_z}\right]$
9. Evaluate $\beta = \sqrt{\sum y_i^2}$

10. Repeat steps 5 to 9 until values converge.

Example 2.6

Example 2.4 will be recalculated using reduced variables.

Solution

$$Z = \sigma \tan\phi - \tau$$

$$\text{or} \quad Z = g(X_1, X_2, C)$$

$$\text{where } X_1 = \sigma;\ X_2 = \phi;\ C = \tau$$

$$\text{Hence} \quad Z = g(X) = X_1.\tan(X_2) - 50$$

i.e. the basic variables are therefore:-

	Mean	s. d.	
X_1	100	20	kN/m^2
X_2	35	5	degrees

$$\text{Now } X_1 = \sigma_1 y_1 + m_1 \quad \text{and } X_2 = \sigma_2 y_2 + m_2$$

$$\text{Hence } Z = g(y) = (y_1\sigma_1 + m_1)\tan(y_2\sigma_2 + m_2) - 50$$

$$g_1' = \frac{\partial Z}{\partial Y}\Big|_{y_1} = \sigma_1 \tan(y_2\sigma_2 + m_2)$$

$$g_2' = \frac{\partial Z}{\partial Y}\Big|_{y_2} = (y_1\sigma_1 + m_1)\sec^2(y_2\sigma_2 + m_2)$$

The first three steps of the iteration procedure have now been carried out. The procedure continues:-

Step 5 With $y_1 = y_2 = 0$ and $= 0$ then $g_1' = 14.004$
and $g_2' = 13.005$

Step 6 $g(y) = 100 \times .7002 - 50 = 20.02$

Step 7 $\sigma_z = \sqrt{14.004^2 + 13.005^2} = 19.112$

Step 8 $y_1 = -\frac{14.004}{19.112}\left[0 + \frac{20.02}{19.112}\right] = -0.767$ and $y_2 = -0.713$

Step 9 $\beta = \sqrt{.767^2 + .713^2} = 1.047$

The procedure is now continued by returning to step 5 and inserting the derived values for y_1, y_2 and β where appropriate.

The complete iteration is set out below.

Iteration	y_1	y_2	β	g(y)
1	0	0	0	20.02
2	-.767	-.713	1.048	1.7416
3	-.890	-.740	1.157	-.0177
4	-.899	-.727	1.156	-.0018
5	-.900	-.725	1.156	-.0000
6	-.900	-.725	1.156	-.0000

The reliability index = 1.156

It can be seen that the value of β is identical to that obtained in terms of the original variables, X_i.

For this two dimensional problem a graphical solution is possible as the failure boundary can be plotted in a similar manner to that of example 2.1. By selecting a range of suitable values for y_1, determining X_1 and calculating the X_2 values, the corresponding set of y_2 values can be obtained.

y_1	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0
X_1 (kN/m^2)	60	70	80	90	100	110	120
X_2 (^{o})	39.8	35.5	32.0	29.1	26.6	24.4	22.6
y_2	0.96	0.1	-0.6	-1.18	-1.68	-2.12	-2.48

The failure boundary is shown plotted in transformed variable space in Fig.2.5. It can be seen that the value of β scales 1.16.

The design point in terms of the standardised variables, y_i, is of course (-0.9, -0.725) but the design point in terms of the original variables can also be obtained :-

$X1 = y_1\sigma_1 + m_1 \quad = -0.9 \times 20 + 100 = 82\text{kN/m}^2$

$X2 = y_2\sigma_2 + m_2 \quad = -0.725 \times 5 + 35 = 31.375^o$

Hence design point is (82kN/m^2, 31.38^o)

Proof that x^{*} corresponds to the maximum probability of failure

The hyperplane is the simplest form of failure surface, Ditlevsen, (1983), and is a surface which has a lineal equation of the form:-

$$a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots\dots\dots\; a_nx_n = 0$$

If the approximation for Z is examined:-

$$Z = \sum^{n}(y_i - y_i^{*})g_i'(y^{*}) = 0$$

it is seen that the equation is of the form just described and that the limit state surface is therefore a hyperplane.

It has been established that, in terms of standardised variables, the design value, y_i^{*}, is:-

$$y_i^{*} = -\alpha_i\beta$$

and that β is the vector, or distance, from the origin of the transformed space to the design point y^{*}.

Hence $\alpha_i = \dfrac{y_i}{\beta} = \dfrac{\text{component of } \beta \text{ on i axis}}{\text{length of } \beta} =$ the direction cosine of β relative to i axis

Provided that the basic varIables that comprise the limit state function have normal distributions then the normal standard density has rotational symmmetry, i.e. all points at the same distance from the origin have equal joint-probability densities. This means that the point on the failure surface that is nearest to the origin must have the highest probability of failure density and is therefore the design point y^{*}.

The fact that y^{*} is the point nearest to the origin can be proved by considering any point, $y^{+} = (y_1^{+}, y_2^{+}, y_3^{+}, \dots\dots y_n^{+})$, on the failure boundary (Z = 0).

The linear approximation to the boundary at point y^+ is:-

$$Z = \sum_{i=1}^{n}(y_i - y_i^+)g_i{}'(y^+) = 0$$

which can be rewritten as:-

$$Z = \sum_{i=1}^{n} y_i g_i{}'(y^+) - \sum_{i=1}^{n} y_i^+ g_i{}'(y^+) = 0$$

This equation represents the hyperplane which is tangential to the failure boundary at point y^+ and is illustrated in Fig.2.5.

The equation may now be rewritten as:-

$$\frac{\sum_{i=1}^{n} y_i g_i{}'(y^+)}{\left|\sum_{i=1}^{n}(g_i{}'(y^+))^2\right|^{1/2}} - \frac{\sum_{i=1}^{n} y_i^+ g_i{}'(y^+)}{\left|\sum_{i=1}^{n}(g_i{}'(y^+))^2\right|^{1/2}} = h$$

where h is the length of the perpendicular from the hyperplane to the origin and where

$$\frac{y_i g_i{}'(y^+)}{\left[\sum_{i=1}^{n}(g_i{}'(y^+))^2\right]^{1/2}} = \cos\theta$$

are its direction cosines.

Comparing this equation with the one for α_i shows that α_i are the direction cosines for the perpendicular to the hyperplane corresponding to a linear approximation to the failure boundary at the point y^*.

This fact has already been established by a study of the equation

$$y_i^* = -\alpha_i\beta$$

and, as the equation $Z = \sum_{i=1}^{n}(y_i - y_i^{*})g_i{}')y^{*}) = 0$ is satisfied, then y^* must also lie on the failure boundary.

The point y^* is therefore the point on the failure boundary which is closest to the origin.

Determination of P_f

Civil engineers are familiar with quality control and therefore readily accept statements such as "out of one thousand concrete slabs manufactured there will be one bad one", or, "the probability of failure of the slabs is 1 in 1000".

This may be why most civil engineers, perhaps unfortunately, are willing to accept the idea that a structure will have a probability of failure value and yet not appreciate the idea of a reliablity index.

Provided that the variables involved have probability distributions that are close to normal and provided that the lineal approximation of the failure surface is realistic then an exact value for P_f can be obtained from the expression:-

$$P_f = \Phi(-\beta)$$

where $\Phi(-\beta)$ is the general symbol for the value of the cumulative probability of Z (for $-\infty$ to $-\beta$). This value can be obtained from tables or from a suitably programmed micro computer or calculator. It is the area under the standardised normal density function and is illustrated in Fig.2.6.

If the variables depart from normal or if the lineal approximation is poor then the P_f value obtained from the above formula is referred to as the "notional, (or nominal), probability of failure".

An important question is whether or not a lineal approximation of the failure surface is accurate enough. Checks can be carried out by Monte Carlo and other simulation methods but a number of better approximations have been developed, Ditlevsen, (1976), Fiessler et al.,(1979), and these generally show very good agreement with first order reliability methods.

These better methods use a quadratic expansion of $g(y) = 0$ and are called "second order reliability methods".

Second order methods have not been used to check the work of this thesis which has been checked by simulation.

It can safely be assumed that, for almost all engineering problems, the lineal approximation of the failure surface will be adequate and it is worth remembering that if there are several variables, of roughly equal weight, the resulting Z function tends to be normal, even when the separate variables are not themselves normal, Benjamin & Cornell (1970).

Nevertheless if it is known that some of the variables involved in the design are non-normal then accuracy of the determined value of P_f is improved if this information is incorporated into the reliability analysis. This can be achieved by a method proposed by Fiessler and Rackwitz, (1976), and described in chapter 6 of this thesis.

However it should be remembered that with geotechnical problems, due to the inevitable lack of statistical information, any P_f values obtained are nominal.

For interest the nominal probability of failure of example 2.4 is

$$P_f = \Phi(-\beta) = \Phi(-1.156) = 0.124 \text{ or some } 12\%.$$

CHAPTER THREE - A SUGGESTED SECOND MOMENT APPROACH FOR GEOTECHNICAL STRUCTURES.

The previous chapters have illustrated how the reliability index, and hence the probability of failure, can be obtained by the application of statistics and probability theory.

Many papers suggesting how these applications can be made have been published and many of them contain mathematical language that the average practising civil engineer last heard in his college days.

There is little doubt that a large amount of the suspicion that most civil engineers have towards the application of probability theory is due to the lack of communication between the erudite mathematicians who write the articles and the practising engineers who could most benefit from them.

For some years now the writer has concerned himself with the possibility of producing a simple method by which the reliability indices of earth retaining structures could be evaluated.

His first thoughts on the formulation of a method were presented for discussion at a workshop on geotechnics in highway design, held at T.R.R.L., Crowthorne, in March 1981.

Later these ideas, suitably refined, were published in the October issue of, "Ground Engineering", (1981) and subsequently discussed at the British Geotechnical Society's symposium on the European Code for foundations, held at the City University in October 1981.

The basis of the method, at that stage, can best be appreciated by describing the suggested procedure for solving the bearing capacity problem of a foundation supported on a granular soil.

The correlation between the angle of friction, ϕ, and the bearing capacity coefficient, N_γ, was established and expressed as an exponential function of ϕ.

With this relationship it became possible to establish an iterative procedure in which the value of N_γ was directly correlated to ϕ.

The method had the merit of being easy to understand but the actual procedure was cumbersome and it was obvious that further improvements were necessary if the method was ever to become a simple means of reliability analysis.

At first glance there seem to be little complications that could prevent the second moment method used for the analysis of structural reliability being modified in order to deal with geotechnical situations.

After all a soil mechanics limit state function contains very few basic variables. For example the strength of a soil structure only involves three variables:- density, unit cohesion and angle of shearing resistance.

However on a closer examination it soon becomes apparent that it is not possible to prepare a simple modification.

The problem is that, although there may be few basic variables in a limit state equation involving soils, there are many terms that are functions of one of these variables, the angle of shearing resistance.

Examples: $K_a = \dfrac{1 - \sin\phi}{1 + \sin\phi}$; $\mu = \tan\phi$; etc.

As an illustration of the complications that can arise consider the relatively simple problem of a concentric column load supported by a square footing, as in example 3.7.

The full limit state function for bearing capacity failure for this example is:-

$$Z = 16.85c[\tan^2(45^{\circ} + \frac{\phi}{2})\exp(\pi\tan\phi) - 1]\cot\phi$$
$$+ 64.8\gamma\tan^2(45^{\circ} + \frac{\phi}{2})\exp(\pi\tan\phi)$$
$$+ 18.66\gamma\tan\phi[\tan^2(45^{\circ} + \frac{\phi}{2})\exp(\pi\tan\phi)$$
$$-27.99\tan\phi - P - 138 = 0$$

In order to use the second moment approach this expression must be differentiated which, although possible, would try the patience of any consulting engineer.

The conversion of the expression into standardised variables also causes more complication with the attendant risk of computational errors.

When it is considered that the average bearing capacity problem will include extra terms such as inclined load factors, eccentricty of loading, etc. it is apparent that there will be many cases in geotechnics where the limit state function will be unmanageable and that some form of simplification must take place if the second moment approach is to be modified for geotechnical work.

The proposed method involves regarding these functions of ϕ as forming a set of independent variables, each with its own expected (or mean) value and its own standard deviation.

Treatment of functions of ϕ

It can be shown that, Benjamin & Cornell, (1970),:-

If $Y = g(X)$ where $X = (X_1, X_2, X_3, \ldots\ldots\ldots X_n)$

$$\text{then } \sigma_Y = \sqrt{\sum_{i=1}^{n} (g_i' \sigma_X)^2}$$

where $g_i' = \frac{\partial Y}{\partial X}\Big|_{m_i}$ i.e. the first derivative of Y with $X = m_i$.

$$\text{If X is a single variable then } \sigma_Y = \sqrt{\left(\frac{\partial Y}{\partial X}\Big|_{m_X} \cdot \sigma_X\right)^2}$$

Example 3.1

The angle of friction of a soil has a mean value of 35° and a standard deviation of 5°. Determine the mean value and standard deviation of $\tan\phi$.

Solution

Let $Y = \tan\phi$.

Then $m_Y = \tan\phi|_{m_\phi} = 0.7002$; $g_\phi' = \frac{\partial Y}{\partial \phi} = \sec^2\phi|_{m_\phi} = 1.490$

$$\sigma_Y = \sqrt{(g_\phi' \sigma_\phi)^2} = \sqrt{(1.49 \times .0873)^2} = 0.1301$$

$\frac{5\pi}{180}$

Example 3.2

The Rankine expression for the coefficient of active earth pressure is:-

$$K_a = \frac{1 - \sin\phi}{1 + \sin\phi}$$

Determine the mean value and standard deviation of K_a for a soil whose angle of friction has a mean value of 33° and a standard deviation of 2°.

Solution

$$\text{Mean value of } K_a = \frac{1 - \sin\phi}{1 + \sin\phi}\bigg|_{m_\phi} = \frac{1 - \sin 33^\circ}{1 + \sin 33^\circ} = 0.2948$$

$$\text{Now } K_a = g(\phi) \quad \text{and} \quad \sigma_{K_a} = \sqrt{\left(\frac{\partial K_a}{\partial \phi}\bigg|_{m_\phi} \cdot \sigma_\phi\right)^2}$$

$$\frac{\partial K_a}{\partial \phi}\bigg|_{m_\phi} = \frac{-2\cos\phi}{(1+\sin\phi)^2}\bigg|_{m_\phi} = \frac{-2\cos 33}{(1+\sin 33)^2} = -\ 0.703$$

$$\text{Standard deviation of } K_a = \sqrt{\left(-0.703 \times \frac{2\pi}{180}\right)^2} = 0.0245$$

Proposed treatment for bearing capacity factors

Meyerhof's equations (1955) for the bearing capacity coefficients N_c and N_q are now generally used in geotechnics as they are recognised as being probably the most satisfactory.

$$N_c = (N_q - 1)\cot\phi \qquad N_q = \tan^2(45^\circ + \frac{\phi}{2})\exp(\pi\tan\phi)$$

Unfortunately there is not the same firmness of opinion about the remaining factor, N_γ. For this text the writer decided to use Hansen's equation (1970):-

$$N_\gamma = 1.5(Nq - 1)\tan\phi$$

$$\text{i.e. } N_\gamma = 1.5\tan\phi[\tan^2(45^\circ + \frac{\phi}{2})\exp(\pi\tan\phi) - 1.5\tan\phi$$

The mean value of N_γ can be quickly found by inserting the mean values of ϕ into the above equation.

The standard deviation of N_γ can be found from the expression:-

$$\sigma_{N_\gamma} = \frac{\partial N_\gamma}{\partial \phi}\bigg|_{m_\phi} \cdot \sigma_\phi$$

involving the differentiation of the equation for N_γ which, although tedious, is relatively simple and leads to the expression:-

$$\frac{\partial N_\gamma}{\partial \phi} = 1.5\tan\phi\left[\frac{2\cos\phi}{(1-\sin\phi)^2}\exp(\pi\tan\phi) + \tan^2(45^\circ + \frac{\phi}{2})\pi\sec^2\phi(\exp\pi\tan\phi)\right]$$

$$+ 1.5\sec^2\phi\left[\tan^2(45^\circ + \frac{\phi}{2})\exp(\pi\tan\phi) - 1\right]$$

Values of N_γ, N_q, N_c and their derivatives are given in appendices I to III at the end of the text.

Example 3.3

A granular soil has an angle of friction with a mean value of 40° and a coefficient of variation of 2.5%. Determine the corresponding mean and standard deviations values for the bearing capacity coefficient, N_γ.

Solution

$$V_\phi = .025 \Rightarrow \sigma_\phi = .025 \times 40 = 1^\circ = .01745 \text{ radians}$$

From Appendix I, for $\phi = 40^\circ$:-

$$\text{Mean } N_\gamma = 79.54; \quad \sigma_{N_\gamma} = 805.05 \times .01745 = 14.051$$

Determination of s.d. values without differentiation

Some of the ϕ functions used in geotechnics are fairly complicated and their differentiation can present problems.

A way around this difficulty is to determine the values of the function for ϕ values one standard deviation on either side of the mean value of ϕ. The standard deviation of the function is then approximately equal to half of the difference between the two values.

Example 3.4

a) Determine the standard deviation of K_a if ϕ has a mean value of 33^{o} and a standard deviation of 2^{o}. (As in example 3.2)

b) Determine the standard deviation of N_γ if ϕ has a mean value of 40^{o} and a standard deviation of 1^{o}. (As in example 3.3)

Solution

a) ϕ values one standard deviation on either side of m_ϕ are 31^{o} and 35^{o}. Inserting these values into the formula for K_a gives K_a values of .3021 and .2710 respectively.

$$\Rightarrow \text{Standard deviation of } K_a = \frac{.3021-.2781}{2} = .0245$$

b) ϕ values one standard deviation on either side of m_o are 39^{o} and 41^{o}. From Appendix I the corresponding N_γ values are 66.76 and 95.05.

$$\Rightarrow \text{Standard deviation of } N_\gamma = \frac{95.05-66.76}{2} = 14.145$$

Example 3.5

Example 2.4 will now be recalculated using the suggested method.

Solution

$$Z = \sigma\tan\phi - \tau$$

$$\text{or} \quad Z = g(X_1, X_2, C)$$

$$\text{where } X_1 = \sigma;\ X_2 = \tan\phi;\ C = \tau$$

$$\text{Hence} \quad Z = g(X) = X_1 . X_2 - 50$$

i.e. the basic variables are therefore:-

		Mean	s. d.
	X_1	100	20
(From Example 3.1)	X_2	.7002	.1301

$$\text{Now } X_1 = \sigma_1 y_1 + m_1 \quad \text{and } X_2 = \sigma_2 y_2 + m_2$$

Hence $g(y) = (y_1\sigma_1 + m_1)(y_2\sigma_2 + m_2) - 50$

$$g_1' = \frac{\partial z}{\partial y}\bigg|_{y_1} = \sigma_1(y_2\sigma_2 + m_2)$$

$$g_2' = \frac{\partial z}{\partial y}\bigg|_{y_2} = \sigma_2(y_1\sigma_1 + m_1)$$

The first three steps of the iteration procedure have now been carried out. The procedure continues:-

Step 5 With $y_1 = y_2 = 0$ and $\beta = 0$ the $g_1' = 14.004$ and $g_2' = 13.01$

Step 6 $g(y) = 100 \times .7002 - 50 = 20.02$

Step 7 $\sigma_z = \sqrt{14.004^2 + 13.01^2} = 19.115$

Step 8 $y_1 = -\frac{14.004}{19.115}\left[0 + \frac{20.02}{19.115}\right] = -0.767$ and $y_2 = -0.713$

Step 9 $\beta = \sqrt{.767^2 + .713^2} = 1.047$

The procedure is now continued by returning to step 5 and inserting the derived values for y_1, y_2 and β where appropriate.

The complete iteration is set out below.

Iteration	y_1	y_2	β	g(y)
1	0	0	0	20.02
2	-.767	-.713	1.048	1.7416
3	-.890	-.740	1.157	-.0177
4	-.899	-.727	1.156	-.0018
5	-.900	-.725	1.156	-.0000
6	-.844	-.759	1.135	-.0000

The reliability index = 1.135

As described in the introduction the nominal probability of failure can be found from the expression $P_f = \Phi(-\beta)$

Hence $P_f = \Phi(-\beta) = \Phi(-1.135) = 0.128$ or 12.8%.

Example 3.5 is an interesting test of the proposed method in that the exact value of β (1.156) is known from the work described in chapter 2.

The proposed method, for this example, is in error by less than 2%.

Example 3.6

A surface reinforced concrete strip foundation, unit weight $24kN/m^3$, is 2m wide, 0.5m thick and will be subjected to a uniform normal pressure, p, of mean value $500kN/m^2$ and coeff. of variation, Vp, = 6%.

The soil is cohesionless, Unit weight:- mean = $18kN/m^3$: V_γ = 5%

Angle of friction:- mean = 40°: V_ϕ = 2.5%

Determine the reliability index against bearing capacity failure.

Solution

The ultimate bearing capacity, q_u, of a surface strip foundation resting on cohesionless soil can be found from the expression:-

$$q_u = 0.5B\gamma N_\gamma$$

For this example the width B can be regarded as constant at 2m and the expression becomes:-

$$q_u = \gamma N_\gamma \quad (=R)$$

S consists of two parts, the applied pressure, p, which can vary, and the weight of the foundation, $0.5 \times 24 = 12kN/m^2$, which can be assumed to be constant.

$$\text{Now } Z = R - S$$

$$= q_u - p - 12$$

$$V_\gamma = 5\% \Rightarrow \sigma_\gamma = 18 \times .05 = 0.9kN/m^3$$

$$V_p = 6\% \Rightarrow \sigma_p = 500 \times .06 = 30kN/m^2$$

and, from example 1.3, Mean $N_\gamma = 79.54$ and $\sigma_{N_\gamma} = 14.051$

There are therefore three variables involved:-

Parameter	Variable	Mean value	Standard deviation
γ (kN/m^3)	X_1	18	0.9
N_γ	X_2	79.54	14.051
p (kN/m^2)	X_3	500	30

$$\text{i.e. } Z = g(X) = X_1.X_2 - X_3 - 12$$

$$= g(y) = (y_1\sigma_1 + m_1)(y_2\sigma_2 + m_2) - (y_3\sigma_3 + m_3) - 12$$

$$g_1' = \sigma_1(y_2\sigma_2 + m_2)$$

$$g_2' = \sigma_2(y_1\sigma_1 + m_1)$$

$$g_3' = -\sigma_3$$

Using the suggested iterative procedure gives the following:-

Iteration	y_1	y_2	y_3	β	g(y)
1	0	0	0	0	919.72
2	-0.941	-3.323	0.394	3.476	39.53
3	-0.439	-3.584	0.446	3.638	-11.586
4	-0.377	-3.546	0.430	3.592	-0.092
5	-0.382	-3.545	0.429	3.591	0.000

Reliablity index = 3.591 (i.e. nominal $P_f = 1.66\text{x}10^{-4}$)

Example 3.7

Details of a reinforced concrete foundation, 3.6x3.6m^2 square, are shown in Fig.3.1.

The foundation is founded at a depth of 5m below the surface of a partially saturated silt which has both a high cohesive and a high frictional strength with the following values:-

$$\text{Cohesion: } m_c = 90\text{kN/m}^2;\quad \sigma_c = 30\text{kN/m}^2$$

$$\text{Angle of friction: } m_\phi = 30^\circ;\quad \sigma_\phi = 3^\circ$$

$$\text{Unit weight: } m_\gamma = 19\text{kN/m}^3;\quad \sigma_\gamma = 0.5\text{kN/m}^3$$

The foundation will support a concentric 1m^2 reinforced concrete column carrying a load of mean value of 25MN and coefficient of variation of 10%.

Determine the reliability index against bearing capacity failure.

Note

Soils of this type are notoriously variable in strength and the drained strength parameters give the best measure of strength, Lumb, (1966).

Solution

Taking unit weight of reinforced concrete as equal to $24kN/m^3$ and assuming that the unit weight of the excavated soil was constant at $18kN/m^3$:-

Wt.of concrete - excavated soil = $(24-18)(3.6^2x1.5 + 1x3.5) = 138kN$

i.e. If P = column load then total load on foundation = P + 138kN

The generally accepted equation for the ultimate load, Q_u, that can be carried by a square foundation, BxB, founded at depth z is:-

$$Q_u = (1.3cN_c + \gamma zN_q + 0.4\gamma BN_\gamma)B^2$$

$$\text{Now } Z = R - S$$

$$= Q_u - (P + 138)$$

The means and standard deviation of the three bearing capacity coefficients, Nc, Nq and N can be found from Appendices I,II and III.

There are therefore six basic variables:-

Variable	Symbol	Mean	s.d.	units
γ	X1	19	0.5	kN/m^3
c	X2	90	30	kN/m^2
N_c	X3	30.14	7.35	
N_q	X4	18.4	6.45	
N_γ	X5	15.07	7.56	
P	X6	25000	2500	kN

$$\text{Hence } Z = (1.3X_2.X_3 + 5X_1.X_4 + 1.44X_1.X_5)3.6^2 - X_6 - 138$$

$$= 16.848X_2.X_3 + 64.8X_1.X_4 + 18.66X_1.X_5 - X_6 - 138$$

and, expressing in reduced variables:-

$$Z = 16.848(y_2\sigma_2 + m_2)(y_3\sigma_3 + m_3) + 64.8(y_1\sigma_1 + m_1)(y_4\sigma_4 + m_4)$$
$$+ 18.66(y_1\sigma_1 + m_1)(y_5\sigma_5 + m_5) - (y_6\sigma_6 + m_6) - 138$$

$$g_1' = \frac{\partial Z}{\partial y_1} = 64.8\sigma_1(y_4\sigma_4 + m_4) + 18.66\;_1(y_5\sigma_5 + m_5)$$

$$g_2' = \frac{\partial Z}{\partial y_2} = 16.848\sigma_2(y_3\sigma_3 + m_3)$$

$$g_3' = \frac{\partial Z}{\partial y_3} = 16.848\sigma_3(y_2\sigma_2 + m_2)$$

$$g_4' = \frac{\partial Z}{\partial y_4} = 64.8\sigma_4(y_1\sigma_1 + m_1)$$

$$g_5' = \frac{\partial Z}{\partial y_5} = 18.66\sigma_5(y_1\sigma_1 + m_1)$$

$$g_6' = \frac{\partial Z}{\partial y_6} = -\sigma_6$$

The iterative procedure for β gives:-

	y_1	y_2	y_3	y_4	y_5	y_6	β	g(y)
1	0	0	0	0	0	0	0	48560
2	-0.08	-1.71	-1.25	-0.89	-0.30	0.28	2.33	7938
3	-0.10	-2.10	-0.95	-1.57	-0.53	0.49	2.88	-1692
4	-0.07	-2.16	-0.62	-1.46	-0.49	0.46	2.76	-390
5	-0.07	-2.25	-0.49	-1.35	-0.45	0.42	2.73	-13.46
6	-0.07	-2.26	-0.47	-1.33	-0.45	0.42	2.73	-2.40
7	-0.07	-2.26	-0.46	-1.33	-0.45	0.42	2.73	-0.43
8	-0.07	-2.26	-0.46	-1.33	-0.45	0.42	2.73	-0.08

Reliability index = 2.73

Example 3.8

Fig.3.3 shows details of a mass stone rubble retaining wall which has a unit weight of 20kN/m^3. The retained soil has a level surface and carries a uniform surcharge, w_s with a mean value of 15kN/m^2 and a standard deviation of 2.5kN/m^2.

The retained soil is granular with the following properties:-

Unit weight: Mean value = 20kN/m^3; s.d. = 1.5kN/m^3

Angle of friction: Mean value = 40°; s.d. = 1.5°

The foundation soil is granular and its angle of friction has a mean value of 40° and a standard deviation of 3°.

The coefficient of friction of the base of the wall and the foundation soil, μ, can be assumed to be equal to the tangent of the angle of friction of the foundation soil.

Assuming that the back of the wall is smooth and using Rankine's formula for K_a, determine the reliability index against sliding.

<u>Solution</u>

The self weight of a structure is not generally treated as a basic variable (see chapter 6) and if we assume that the weight of the wall, W, is constant then:-

$$W = 6 \times 20 \times \frac{(2+1)}{2} = 180\text{kN}$$

S = Total horizontal thrust from soil on to back of wall

$$= 6.K_a.w_s + \frac{1}{2}.\gamma.K_a.6^2 = 6K_a.w_s + 18.\gamma.K_a$$

$$R = \text{Sliding resistance} = W.\mu = 180\mu$$

$$\Rightarrow Z = 180\mu - 6K_a.w_s - 18K_a\gamma$$

There are four basic variables:- γ, K_a, w_s and μ and, by using the methods of the previous examples, their means and standard deviations can quickly be found.

Variable	Symbol	Mean	s.d.	Units
γ	X_1	20	1.5	kN/m^3
K_a	X_2	.2174	.0149	
w_s	X_3	15	2.5	kN/m^2
μ	X_4	.8391	.0893	

Hence $Z = 180X_4 - 6X_2.X_3 - 18X_1.X_2$

$$= 180(y_4\sigma_4+m_4) - 6(y_2\sigma_2+m_2)(y_3\sigma_3+m_3) - 18(y_1\sigma_1+m_1)(y_2\sigma_2+m_2)$$

$$g_1' = -18\sigma_1(y_2\sigma_2 + m_2)$$

$$g_2' = -6\sigma_2(y_3\sigma_3 + m_3) - 18\sigma_2(y_1\sigma_1 + m_1)$$
$$g_3' = -6\sigma_3(y_2\sigma_2 + m_2)$$
$$g_4' = 180\sigma_4$$

The iterative procedure for β gives:-

Iteration	y_1	y_2	y_3	y_4	β	g(y)
1	0	0	0	0	0	53.21
2	0.90	1.02	0.50	-2.46	2.85	-0.48
3	0.93	1.07	0.52	-2.39	2.83	-0.03
4	0.94	1.07	0.52	-2.39	2.82	-0.00
5	0.94	1.07	0.52	-2.39	2.82	0.00

The reliability index = 2.82

<u>Checks on the proposed method</u>

The reliability indices obtained in examples 3.6 and 3.7 were checked by the full correlation method, Smith, (1981) and by Monte Carlo simulation.

A Monte Carlo simulation is simply a technique in which an output of random numbers is related to an assumed probability distribution (generally normal but other distributions may be used) so that a set of probable values for the basic variables of the Z function is obtained after an iteration.

With these values the corresponding factor of safety, F = R/S, for that particular iteration be obtained.

If enough iterations are carried out then a histogram, and hence a frequency curve, of F can be obtained, such as illustrated in Fig.3.2.

The probability of failure can be expressed as the distance from the mean value of F to the failure point (the point where F = 1.0).

If this distance is expressed as a multiple of σ_F then we obtain the value of β.

$$\beta = \frac{m_F - 1}{\sigma_F}$$

In order to remove the possibility of outliers and to make certain that ϕ values did not reduce below ϕ_{cv} values the generated maximum and minimum values of the basic variables were set at the values that had less than a 5% probability of occurrence.

i.e. Maximum value = mean value + 1.645 s.d
Minimum value = mean value - 1.645 s.d.

Each basic variable is treated as being independent, i.e. separate values of ϕ were generated for the determination of each of the three bearing capacity coefficients in example 3.7.

Grigoriu, (1983) maintains that if $P_f = 10^{-4}$ then the number of iterations should be in the order of 10^5. In fact in the last four examples, and in the others of later chapters that were also checked, a stable value for β was established after some 5000 iterations. In most cases 10,000 iterations were carried out.

The values of β obtained by the three methods for the four examples are shown below.

	Full correlation	Proposed Method	Monte Carlo
Eg.3.5	1.16	1.14	1.15
Eg.3.6	5.49	3.59	3.80
Eg.3.7	2.60	2.73	2.44
Eg.3.8	2.82	2.82	2.64

The value of 2.73 obtained for β by the proposed method in example 3.7 is slightly high. A more accurate value is 2.65 which is obtained when the distributions of the three bearing capacity coefficients are allowed for. (See example 7.1).

The value of 5.49 obtained for example 3.6 by the full correlation method seems, intuitively, to be unreasonably high and yet, although several attempts were made, no computational error could be found.

It may be that the Z function, formulated with full correlation, does not truly represent the limit state surface g(y) = 0.

An investigation of this phenomenon, by some form of mathematical treatment, will prove of interest but, even at this stage it can be said that there are promising indications that the proposed method agrees very well with the other two approaches. Where there is a divergence the value obtained by the method appears to be the more reasonable and, in any case, is on the safe side.

CHAPTER FOUR - RELIABILITY ANALYSES OF GEOTECHNICAL STRUCTURES

The proposed reliability analysis method will now be used to determine the nominal probability of failure value for two simple geotechnical structures, a retaining wall and a cutting. The word 'simple' has been included as it has been assumed that the soils in both of the following examples are homogeneous.

Example 4.1

Details of a retaining wall are given in Fig.4.1.

The relevant properties of the two soils can be assumed to be:-

Fill material - a granular soil with no cohesion.

Angle of friction, ϕ_1 :- Mean = 35°; s.d. = 1°

Unit weight, γ_1:- Mean = $19kN/m^3$; s.d. = $1kN/m^3$

Foundation soil - a granular soil with no cohesion.

Angle of friction, ϕ_2 :- Mean = 35°; s.d. = 1.5°

Unit weight, γ_2:- Mean = $19kN/m^3$; s.d. = $1.5kN/m^3$

The coefficient of friction, μ, between the wall base and the foundation soil can be taken as equal to $\tan\phi_2$.

All variables may be assumed to have normal distributions and the dimensions of the structure and the unit weight of the concrete ($24kN/m^3$) taken as constant.

The coefficient of active earth pressure, K_a, can be taken as equal to the Coulomb value, (See Smith, 1982):-

$$K_a = \left(\frac{\operatorname{cosec}\psi \sin(\psi - \phi)}{\sqrt{\sin(\psi + \delta)} + \sqrt{\dfrac{\sin(\phi + \delta)\sin(\phi - i)}{\sin(\psi - i)}}} \right)^2$$

where ψ = angle of back of wall to the horizontal, δ = angle of wall friction, i = angle of inclination of surface of retained soil to the horizontal and ϕ = angle of friction of the retained soil.

Determine the reliability index for each failure mode given that the probability of failure by rotational slip is less than 10^{-8} and may be considered negligible.

Solution

The investigation of the stability of a retaining wall involves the examination of four modes of failure:-

i) Sliding
ii) Overturning
iii) Bearing capacity
iv) Rotational slip

i - Sliding

The first step is to determine the limit state function for sliding.

Height of AB = $5.5 + \frac{2.5}{3} = 6.333\text{m}$

Thrust from soil, P_a, $= 0.5K_a\gamma_1 6.333^2 = 20.055\gamma_1 K_a$

$$P_{aH} = P_a\cos\phi_1 \quad (= S); \quad P_{aV} = P_a\sin\phi_1$$

R = the resistance to sliding = μ x Vertical reaction = μR_v

where R_v = Weight of wall + soil on heel + P_{aV}

$= 24(5\text{x}.375 + 4\text{x}.5) + 13.543\gamma_1 + 20.055K_a\gamma_1\sin\phi_1$

$= 93 + 13.543\gamma_1 + 20.055K_a\gamma_1\sin\phi_1$

Z = R - S

$= (93 + 13.543\gamma_1 + 20.055K_a\gamma_1\sin\phi_1) - 20.055K_a\gamma_1\cos\phi_1$

The basic variables are therefore γ_1, K_a, $\cos\phi_1$, $\sin\phi_1$ and μ.

If we designate them as X_1 to X_5 respectively then:-

$$Z = g(X) = (93 + 13.543X_1 + 20.055X_1.X_2.X_4)X_5 - 20.055X_1.X_2.X_3$$

The means and standard deviations of these variables can be found by the method suggested in chapter 3.

Variable	Mean	Standard deviation
X_1	19	1
X_2	.3252	.0141
X_3	.8192	.0100
X_4	.5735	.0143
X_5	.7002	.0390

Expressing the limit state function in standardised variables:-

$Z = g(y)$

$$= [93+13.543(y_1\sigma_1+m_1)+20.055(y_1\sigma_1+m_1)(y_2\sigma_2+m_2)(y_4\sigma_4+m_4)](y_5\sigma_5+m_5)$$
$$- 20.055(y_1\sigma_1+m_1)(y_2\sigma_2+m_2)(y_3\sigma_3+m_3)$$

The determination of the various derivatives of g(y) is a straightforward procedure. One example will be given.

$$g_4' = 20.055\sigma_4(y_1\sigma_1+m_1)(y_2\sigma_2+m_2)(y_5\sigma_5+m_5)$$

The iterative procedure to determine β gives:-

	y_1	y_2	y_3	y_4	y_5	β	g(y)
1	0	0	0	0	0	0	193.52
2	-4.04	1.34	0.74	-0.74	-9.82	10.76	26.86
3	0.16	2.46	0.93	-0.42	-12.38	12.67	-20.77
4	1.41	2.51	0.91	-0.28	-11.06	11.47	-0.29
5	0.78	2.49	0.93	-0.36	-11.12	11.45	-0.31
6	0.82	2.44	0.92	-0.35	-11.10	11.43	-0.00
7	0.81	2.44	0.92	-0.35	-11.10	11.43	-0.00
8	0.81	2.44	0.92	-0.35	-11.10	11.43	-0.00

Reliability index = 11.43

Note - Maximum and minimum values of β

There are obviously occasions when a particular limit function has a resistance, R, far in excess of the disturbing load, S. In such a situation Z can only equal zero if β is extremely large and at least some of the resistance variables achieve unrealistically low, possibly negative, values.

For large and important structures the highest acceptable level of P_f is about 10^{-8}, the acceptable risk of a design or constructional fault which could lead to major structural damage and loss of life.

The acceptable risk for a catastrophy such as a nuclear explosion, is about 10^{-10} but such a risk is outwith normal structural engineering considerations.

A Pf value of 10^{-8} corresponds to a β value of 5.6. Hence, in order to avoid negative variables, etc. it is sensible to accept that the risk of failure is negligible when the value of Z is so large that β must exceed 5.6 for it to be reduced to zero.

Hence, for the sliding failure mode of this example it is more sensible to say that the risk of sliding failure is negligible rather than attempt to obtain a P_f value.

Acceptable maximum P_f values for geotechnical engineering have been suggested by Meyerhof, (1982):-

	Maximum P_f	Corresponding β value
Earthworks	10^{-2}	2.33
Earth retaining structures	10^{-3}	3.10
Offshore foundations	10^{-3}	3.10
Onshore foundations	10^{-4}	3.80

ii - Overturning

Taking moments about point C, the toe of the wall, establishes:-

Resistive moments (R)

M due to wall stem and slab = 154.75kNm

M due to soil on heel = $37.68\gamma_1$

M due to P_{aV} = $80.22K_a\gamma_1\sin\phi_1$

Disturbing moment (S)

$S = 42.34K_a\gamma_1\cos\phi_1$ (assuming that P_{aH} acts at 0.333AB above base of slab)

$Z = R - S$

$= 154.75 + 37.68\gamma_1 + 80.22K_a\gamma_1\sin\phi_1 - 42.34K_a\gamma_1\cos\phi_1$

$$g(y) = 154.75 + 37.68(y_1\sigma_1+m_1) + 80.22(y_1\sigma_1+m_1)(y_2\sigma_2+m_2)(y_4\sigma_4+m_4) - 42.34(y_1\sigma_1+m_1)(y_2\sigma_2+m_2)(y_3\sigma_3+m_3)$$

where $X_1 = \gamma_1$; $X_2 = K_a$; $X_3 = \cos\phi_1$; $X_4 = \sin\phi_1$

The iterative procedure gives:-

Iteration	y_1	y_2	y_3	y_4	β	g(y)
1	0	0	0	0	0	940.62
2	-21.89	-1.61	1.38	-3.75	22.3	40.11
3	-23.32	0.15	-0.22	0.59	23.3	-24.96
4	-22.70	0.40	-0.33	0.88	22.7	-0.06
5	-22.71	0.35	-0.28	0.76	22.7	-0.02
6	-22.71	0.35	-0.28	0.76	22.7	-0.00

The risk of overturning failure is negligible, as $\beta > 5.6$.

iii - Bearing Capacity

For a surface strip footing the ultimate bearing pressure, q_u, is given by the expression:-

$$q_u = 0.5B\gamma N_\gamma i_\gamma$$

where B = width of foundation
γ = unit weight of supporting soil
N_γ = bearing capacity coefficient
i_γ = inclined load factor

The ultimate vertical load, Q_u, that can act of the foundation is found from the expression:-

$$Q_u = B'q_u$$

where $B' = (B - 2e)$

and e = eccentricity of R_v (the vertical reaction)

Hence $R = Q_u = 0.5(B - 2e)B\gamma_2 N_\gamma i_\gamma = 2(4 - 2e)\gamma_2 N_\gamma i_\gamma$

and $S = R_v = 93 + 13.543\gamma_1 + 20.055K_a\gamma_1\sin\phi_1$

$\Rightarrow Z = 2(4 - 2e)\gamma_2 N_\gamma i_\gamma - 93 - 13.543\gamma_1 - 20.055K_a\gamma_1\sin\phi_1$

The basic variables are γ_1, γ_2, K_a, $\sin\phi_1$, N_γ, i_γ and e. The mean and standard deviations of the first four have already been determined.

The bearing capacity coefficient $N\gamma$

The treatment of this factor has been discussed in chapter 3. For $\phi = 35^\circ$ the mean value of $N\gamma$ is 33.92 (From Appendix I). The value of $\partial N\gamma/\partial\phi$ is 321.31 and $\sigma_{N\gamma}$ is therefore 8.41.

The inclined load factor $i\gamma$

Various expressions have been proposed for i_γ and the one used in this text is that suggested by Sokolovski, (1960):-

$$i_\gamma = \left(1 - \frac{P_{aH}}{R_v}\right)^3$$

Now P_{aH} = Horizontal thrust = $20.055K_a\gamma_1\cos\phi_1$

$Rv = 93 + 13.542\gamma_1 + 20.055K_a\gamma_1\sin\phi_1$

$$\Rightarrow \quad i_\gamma = \left(1 - \frac{20.055K_a\gamma_1\cos\phi_1}{93 + 13.543\gamma_1 + 20.055K_a\gamma_1\sin\phi_1}\right)^3$$

Mean value for i_γ (for $\phi_1 = 35^\circ$ and $\gamma_1 = 19\text{kN/m}^3$) = 0.4378

It can easily been shown that i_γ is, for all practical purposes, only sensitive to changes in the value of ϕ. Hence, assuming that γ_1 is constant at 19kN/m^3 then $\sigma_{i\gamma}$ is found (by the approximate method) to be = 0.0218.

The eccentricity - e

$$e = \left|\frac{M}{R_v} - \frac{B}{2}\right| = \left|\frac{M}{R_v} - 2\right|$$

Taking moments about A, the heel of the wall:-

$$e = \left| \frac{217.25 + 16.49\gamma_1 + 42.306K_a\gamma_1\cos\phi_1}{93 + 13.543\gamma_1 + 20.055K_a\gamma_1\sin\phi_1} - 2 \right|$$

Mean value of e (when $\phi_1 = 35^\circ$ and $\gamma_1 = 19\text{kN/m}^3$) = 0.2325

By the approximate method, assuming γ_1 constant at 19kN/m^3, the standard deviation of e works out at 0.0227.

Designating the variables as X_1 to X_7:-

Variable	Mean	s.d
$\gamma_1 = X_1$	19	1
$\gamma_2 = X_2$	19	1.5
$K_a = X_3$	.3252	.0141
$\sin\phi_1 = X_4$	.5735	.0143
$N_\gamma = X_5$	33.92	8.41
$i_\gamma = X_6$	.4378	.0218
$e = X_7$	.2352	.0227

$$Z = g(X) = 2(4 - 2X_7)X_2.X_5.X_6 - 93 - 13.543X_1 - 20.055X_1.X_3.X_4$$

$$g(y) = 2[4-2(y_7\sigma_7+m_7)](y_2\sigma_2+m_2)(y_5\sigma_5+m_5)(y_6\sigma_6+m_6) - 93$$
$$-13.543(y_1\sigma_1+m_1) - 20.055(y_1\sigma_1+m_1)(y_3\sigma_3+m_3)(y_4\sigma_4+m_4)$$

Iteration gives:-

Iteration	y_1	y_2	y_3	y_4	y_5	y_6	y_7	β	g(y)
1	0	0	0	0	0	0	0	0	1562
2	0.10	-0.88	0.17	0.01	-2.78	-0.56	0.07	2.97	135
3	0.13	-0.35	0.02	0.01	-3.24	-0.21	0.03	3.27	-50.0
4	0.12	-0.20	0.02	0.01	-3.16	-0.13	0.02	3.17	-1.13
5	0.11	-0.22	0.02	0.01	-3.15	-0.14	0.02	3.16	-0.05
6	0.11	-0.22	0.02	0.11	-3.15	-0.14	0.02	3.16	-0.00

The reliablity index is 3.16 and the nominal P_f = 0.0008

Comparison between F and P_f values

The factors of safety, based on the mean values of the variables, are:-

Sliding	1.94
Overturning	5.39
Bearing capacity	4.73

The probabilities of failure are:-

Sliding	- negligible
Overturning	- negligible
Bearing capacity	- 8.0×10^{-4}

It is interesting to note that whilst the factors of safety indicate that the wall is safer in bearing capacity than in sliding the actual risk of sliding is negligible whilst the risk of bearing capacity failure, although acceptable, is the greatest risk of failure of the wall.

Monte Carlo simulations gave the following values for β :-

Sliding	>5.6
Overturning	>5.6
Bearing	3.14

Example 4.2

The method will now be used for the reliability analysis of a cutting in a homogenous c-ϕ soil.

The dimensions of the cutting are shown in Fig.4.2 and the properties of the basic variables can be assumed to be:-

	Mean	s.d.	units
Unit weight, γ	19	1.0	kN/m^3
Angle of shearing resistance, ϕ'	5	1.0	degrees
Unit cohesion, c'	45	20	kN/m^2
Pore pressure ration, r_u	0.4	0.1	

The distributions of all of these variables are assumed to be normal.

With the aid of a computer programme various slip circles were analysed and it was eventually found that the circle intersecting a point 2m in front of the toe was the critical slip circle and, with mean values, gave a factor of safety, F, of 2.30, based on Bishop's conventional formula, 1955), Fig.4.2.

Contours of F, corresponding to the centres of various slip circles, are shown in Fig.4.3A.

It is interesting to note that, in spite of ϕ having a mean value of 5, the minimum value obtained for F for slip circles passing through the toe was 2.37. This is possibly because the conventional, rather than the rigorous formula, was used. The rigorous formula will be adopted for future work but, as the reader will appreciate, the actual position of the slip circle does not detract from this illustration of the proposed method.

Bishop's conventional formula for the factor of safety is:-

$$F = \frac{1}{\sum W\sin\alpha} \sum [c'l + W(\cos\alpha - r_u \sec\alpha)\tan\phi']$$

$$\text{now } F = \frac{\text{restraining moment}}{\text{disturbing moment}} = \frac{R}{S}$$

Hence the limit state equation can be written as:-

$$Z = [c'l + W(\cos\alpha - r_u \sec\alpha)\tan\phi' - \Sigma\, W\sin\alpha$$

$$\text{or } Z = [c'l \;+ \gamma(FA - r_u FR)]\tan\phi' - \gamma D$$

where $FA = \Sigma FA_i$ and $FR = \Sigma FR_i$

with $FA_i = A_i \cos\alpha_i$ and $FR_i = A_i r_u \sec\alpha_i$

and where $l = R\theta$ (R = radius) and $D = \sum_{i=1}^{n} A_i \sin\alpha_i$

Note

Unfortunately there is another clash of symbols here. Bishop used α to designate the angle between the reactive force acting at the mid-point of the base of a slice and the vertical. It is in this sense that α is used here, not as a sensitivity coefficient.

If, in any slice i, the value of $\cos_i\alpha - r_u \sec_i\alpha < 0$ then this value must be put equal to 0. For computational purposes this was achieved by putting both FA_i and FR_i equal to 0.

Using the method described below it was possible to analyse various slip circles and to establish the β contours shown in Fig.4.3B. It was found, that for this example at least, the centre of the slip circle corresponding to minimum β was not in quite the same position as the centre of the circle for minimum F. The coordinates for minimum β were (6.5,8.5) and for minimum F were (7,8.5).
The first step in the iterative procedure was to establish the various constants of the limit state equation.

As the constant FR involves the value of the pore pressure ratio

r_u this term had to be recalculated after each iteration.

With r_u at its initial value of 0.4 the various constants obtained were as set out below:-

Slice (i)	z(m)	b(m)	Area (A_i)	α_i	$FA_i = A_i \cos\alpha_i$	$FR_i = A_i r_u \sec\alpha_i$
1	0.39	1.0	0.39	-36.18	0.31	0.19
2	1.04	1.0	1.04	-29.97	0.90	0.48
3	2.46	1.75	4.31	-22.00	3.99	1.86
4	4.50	1.75	7.88	-12.46	7.69	3.23
5	6.24	1.75	10.92	-3.25	10.91	4.38
6	7.70	1.75	13.84	5.85	13.41	5.42
7	7.69	4.36	33.53	22.31	31.02	14.50
8	4.45	4.36	19.40	50.88	0.00	0.00

The basic variables are therefore γ, ϕ', c' and r_u which are designated as X_1, X_2, X_3 and X_4.

The complete iteration, for the critical slip circle, is set out below:-

Iteration	y_1	y_2	y_3	y_4	β	g(y)
1	0	0	0	0	0	604.59
2	.064	-.038	-1.35	0.038	1.35	0.001
3	.064	-.038	-1.35	0.037	1.35	0.000
4	.064	-.038	-1.35	0.037	1.35	0.000

The reliability index = 1.35

Monte Carlo simulation, after 10,000 iterations, gave $\beta = 1.62$.

Note

Although there are plenty of articles that deal with the reliability analysis of earth slopes there are few that give the dimensions of a slope together with the soil properties and then actually determine a value for β.

The closest that the writer has found is the paper by Chowdhury and A-Grivas, (1982), where the reliability of an earth slope is analysed in terms of progressive failure.

The slope was divided into nine slices and the probability of failure of each slice was found by the authors to be:-

Slice	P_f	Slice	P_f
1	6.95×10^{-3}	5	3.79×10^{-3}
2	1.00×10^{-3}	6	2.33×10^{-2}
3	4.19×10^{-4}	7	7.08×10^{-2}
4	8.45×10^{-4}	8	1.36×10^{-1}
		9	4.50×10^{-2}

It is of interest to note that analysing the same slope by the method suggested here gives an overall reliability index of 2.07 which indicates a nominal probability of failure of some 1.9×10^{-2}.

CHAPTER FIVE - SOIL DISTRIBUTIONS

As has been indicated in the earlier chapters, the forms of the distributions of the basic variables in the limit state function have an effect on the accuracy of β and hence on the prediction of the probability of failure.

Generally there will be little statistical information available to the soils engineer whereby he can accurately determine the various distributions of the soil parameters that he is interested in.

If he has worked for some time in the area he may have some a priori knowledge, in the form of experience with similar soils or even results from similar work on a nearby site. With such knowledge he should be able to at least make meaningful assumptions as to the soil distributions.

For most practical situations the variability of a soil parameter can be adequately described by one of three distributions:-

i) The normal distribution
ii) The lognormal distribution
iii) The beta distribution

For the benefit of readers unfamiliar with these distributions a few brief notes and examples now follow.

The normal distribution

This is the most important of the three distributions. Its probability density function, for a variable X, has the equation:-

$$f_X(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp[-(x-m_x)^2/2\sigma_x^2] \qquad -\infty \leq x \leq \infty$$

Example 5.1

The undrained shear strength of a clay has a mean value of

44.85kN/m^2 and a standard deviation of 6.56kN/m^2.

Fit these values to a normal distribution.

Solution

Substituting for m_x and σ_x in the formula gives:-

$$f_X(x) = \frac{1}{16.44} \exp[-(x-44.85)^2/86.07]$$

By inserting values for x into the above formula the corresponding values of $f_X(x)$ can be obtained.

Eg. For x = 30 $f_X(x) = 0.0047$

If values for x, covering an appropriate range, are inserted into the equation the normal distribution is obtained, (Fig.5.1).

Example 5.2

Using the results of example 5.1 determine the probability that the value of X will lie within the range 40 to 50KN/m^2.

Solution

Values of x and the corresponding values of $f_X(x)$ over the range x = 40 to x = 50kN/m^2 are set out below.

x	40	41	42	43	44	45	46	47	48	49	50
	.0463	.0512	.0553	.0584	.0603	.0608	.0599	.0576	.0542	.0498	.0447

By Simpson's rule, area under curve = 0.554. i.e. Probability = 0.554

Alternative solution - using standardised variables

The properties of the normal distribution are well documented and tables of values of $f_X(x)$, etc. for this distribution are readily available. In order to use this information it is first necessary to standardise the x values into a new set of z values:-

$$z = \frac{x - m_x}{\sigma_x}$$

The pdf of Z is known as the standardised normal density function and has the property that its mean = 0 and its variance = 1. Because of this the pdf is often symbolised as N(0,1) whereas the pdf of the normal distribution of example 5.1 is written as N(44.85,6.56).

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \qquad \text{(See Fig.5.2)}$$

Values of both $f_Z(z)$ and of $F_Z(z)$, (the cumulative density function, symbol cdf), are listed in statistical tables, etc.

$$z_{40} = (40-44.85)/6.56 = -0.7393$$

And $P[Z = 40] = F_Z(40)$ .... From tables, etc. $F_Z(40) = 0.238$

$$z_{50} = (50-44.85)/6.56 = 0.7851$$

And $P[Z = 50] = F_Z(50)$..... From tables, etc. $F_Z(50) = 0.784$

Hence the probability that X will lie between 40 and $50kN/m^2$ is:-

$$0.784 - 0.230 = 0.554$$

The lognormal distribution

A variable is said to have a lognormal (or logarithmicnormal) distribution when the logarithms of its values are normally distributed.

We can consider this distribution mathematically if we think in terms of two variables, X and Y, such that Y = exp(X)

If the relationship between the values of two variables is of the form:-

$$Y = g(X)$$

and if y increases as x increases and if there is only a single value of y corresponding to each value of x, and viceversa, then we say that y is a monotonically increasing function with x.

Generally if we know the function Y = g(X) we can find the inverse function:-

$$X = g^{-1}(Y)$$

In the case of a monotonically increasing function we can solve directly for the cdf of Y as the probability that Y is equal to or less than a particular value, y, must be equal to the probability that X is equal to or less then the value x, where x is the value corresponding to y, .i.e. $g^{-1}(y)$.

Hence $F_Y(y) = P[Y \leq y] = P[X \leq x] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$

Now the pdf of Y can be obtained by differentiating its cdf:-

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}[F_X(g^{-1}(y))] = \frac{d}{dy}\int_{-\infty}^{g^{-1}(y)} f_X(x)dx$$

which simplifies to:-

$$\text{i.e. } f_Y(y) = \frac{dg}{dy}^{-1}(y)f_X(g^{-1}(y))$$

$$= \frac{dx}{dy} f_X(x) \qquad \text{..........(A)}$$

In the case of a lognormal distribution:-

$$X = \ln Y$$

$$Y = \exp(X)$$

Hence we can say that $Y = g(X) = \exp(X)$

$X = g^{-1}(Y) = \ln Y$ (normally distributed)

and $\frac{dx}{dy} = \frac{1}{y}$

As X is normally distributed we can write down:-

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp[-\frac{1}{2}\left(\frac{x - m_x}{\sigma_X}\right)^2 \qquad -\infty \leq x \leq \infty$$

And, with the relationship of expression (A) and substituting gives:-

$$f_Y(y) = \frac{1}{y\sigma_X\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(y) - m_X}{\sigma_X}\right)^2\right] \qquad y \geq 0$$

Note that X cannot have a negative value as the expression $X = \ln(-y)$ is meaningless.

It will be seen that the expression just derived for $f_Y(y)$ is in terms of m_X and σ_X, the mean and standard deviation of X, the logarithms of the values of Y. The expression can be improved if we make a substitution for $(\ln(y) - m_X)$.

In order to enable us to do this we must think in terms of the medians of X and of Y.

The median of a variable, given the symbol $\breve{m}$, is simply the middle value of the distribution when the values are placed in ranking order. If there are an even number of values then the median value is taken to be the average of the two central values.

It is obvious that the probability of a variable X having a value equal to or less than its median, $F_X(\breve{m}_X)$, = 0.5.

Now for any value y the corresponding value of $x = \ln(y)$

$$\text{Hence } P[Y \leq \breve{m}_Y] = P[X \leq \ln(\breve{m}_Y)] = 0.5$$

$$\text{and } P[X \leq \breve{m}_X] = 0.5$$

$$\text{Hence } \ln(\breve{m}_Y) = \breve{m}_X$$

For a normal, or indeed any symmetrical, distribution $\breve{m}_X = m_X$

$$\Rightarrow \ln(\breve{m}_Y) = m_X$$

$$\text{Hence } (\ln(y) - m_X) = \ln(y) - \ln(\breve{m}_y)$$

Substituting in the expression for $f_Y(y)$:-

$$f_Y(y) = \frac{1}{y\sigma_X\sqrt{2\pi}} \exp\left[-0.5\left(\frac{\ln(y) - \ln(\breve{m}_Y)}{\sigma_X}\right)^2\right] \qquad y \geq 0$$

Note It can be shown that:-

$$\sigma_X^2 = \ln(V_Y^2 + 1) \quad \text{where } V_Y = \text{coeff. of variation of } Y = \frac{\sigma_Y}{m_Y}$$

$$\breve{m}_Y = m_Y \exp(-0.5\sigma_X)^2$$

Example 5.3

Using the same mean and standard deviation values of example 5.1 determine the probability that X will lie between 40 and 50kN/m^2, assuming a lognormal distribution.

Solution

It is perhaps useful to keep the symbol for the variable as Y to serve as a reminder that the variable is related.

$$V_Y = \frac{\sigma_Y}{m_Y} = \frac{6.56}{44.85} \doteq 0.1463$$

$$\sigma_X^2 = \ln(0.1463^2 + 1) = 0.212 \quad \Rightarrow \quad \sigma_X = 0.1455$$

$$\breve{m}_Y = 44.85\exp(-0.5 \times 0.0212) = 44.38\text{kN/m}^2$$

$$\Rightarrow f_Y(y) = \frac{1}{0.1445\sqrt{2\pi}\, y} \exp\left| -0.5 \frac{\ln(y) - \ln 44.38}{0.1455}^2 \right|$$

The appropriate values of $f_Y(y)$ corresponding to y are:-

y	40	41	42	43	44	45	46	47	48	49	50
	.0531	.0577	.0808	.0623	.0622	.0607	.0578	.0540	.0494	.0444	.0392

and, using Simpson's rule, area under pdf curve = 0.556

Alternative solution - using normal distribution information

We have already established, for a lognormal distribution, the relationship between Y and X:-

$$f_Y(y) = \frac{dx}{dy} f_X(x)$$

If we express $f_Y(y)$ in terms of its standardised variable then

$$f_Y(y) = \frac{dz}{dy} f_Z(z) \qquad \text{where } z = \frac{\ln(y) - \ln(\breve{m}_Y)}{\sigma_X}$$

$$\text{Now } \frac{dz}{dy} = \frac{1}{y} \cdot \frac{1}{\sigma_X}$$

$$\Rightarrow \quad f_Y(y) = \frac{f_Z(z)}{y\sigma_X}$$

Tables for the normal distribution can now be used to determine values of $f_Y(y)$. For example, for $f_Y(45)$:-

$z = (\ln 45 - 3.7928)/0.1455 = 0.01386/0.1455 = 0.0953$

From the tables, for $z = 0.0953$, $f_Z(z) = 0.3972$

$$\Rightarrow fY(y) = \frac{0.3972}{45 \times .1455} = 0.0607$$

However the main advantage of using Z is in the determination of $F_Y(y)$ values, as $F_Y(y) = F_Z(z)$.

$z_{40} = -0.7142$ .... From tables, etc. $F_Z(-.7142) = 0.2375$

$z_{50} = 0.2375$ ..... From tables, etc. $F_Z(.2375) = 0.7932$

Hence $P[40 \le Y \le 50] = 0.7932 - 0.2375 = 0.556$

<u>The beta distribution</u>

The pdf of this distribution, in its general form, is:-

$$f_X(x) = \frac{1}{B(b-a)^{t-1}} (x-a)^{r-1}(b-x)^{t-r-1} \qquad a \le X \le b$$

where a and b = the minimum and maximum values of X respectively

r and t = numerical constants related to the mean and variance of X.

B = the normalising constant.

The beta distribution is usually symbolised as BT(r,t) and the range of shapes of distribution curves that it can represent is remarkable varying from rectangles to symmetrical or asymmetrical curves.

Necessary information

The mathematics behind this distribution will not be examined but if the reader wishes to use the distribution he will require the following information.

$$m_X = a + \frac{r}{t}(b-a); \qquad \sigma_X^2 = \frac{(b-a)^2 r(t-r)}{t^2(r+1)}$$

Provided that r and t are integers B can be found from the formula:-

$$B = \frac{(r-1)!(t-r-1)!}{(t-1)!}$$

If r and t are not whole numbers approximation can be attempted but it is probably best to determine B by using the fact that the total area under the pdf curve is equal to 1.0

The procedure is to determine a suitable number of values of $f_X(x)$ over the range of X values using the formula, with B put equal to 1.0.

With these values and by Simpsons rule the area under the curve (with B = 1.0) can be obtained. This value of the area will be the true value of B.

The method is simple to understand but involves a lot of tedious work which is best computerised.

Values of a and b

The minimum and maxiumum values of X, a and b, must be known if $f_X(x)$ is to be determined and suggested approximations for soil parameters are given later in this chapter.

Example 5.4

A series of measurements on the angle of shearing resistance of a sand determined that the parameter had a mean value of 35°, a standard deviation of 3° and that its minimum and maximum values were 28° and 42° respectively.

Fit the results to a symmetrical beta distribution.

Solution

$$m_{\phi} = 35^{\circ}; \ \sigma_{\phi} = 3^{\circ}$$

$$\Rightarrow \quad 35 = 28 + \frac{r}{t}(42 - 28) = 28 + 14\frac{r}{t}$$

$$\Rightarrow \quad r = 0.5t \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (A)$$

$$\sigma_{\phi}^{2} = 3^{2} = 14^{2} \times \frac{0.5(t - .5)}{t^{2}(t + 1)}$$

Hence $t = 4.44$ and, from (A), $r = 2.22$

If we assume $t = 4$ and $r = 2$ then $B = \frac{(2-1)!(4-2-1)!}{(4-1)!} = \frac{1}{3!} = 0.1667$

This figure would have to be used unless there was access to a microcomputer when a more accurate figure could be obtained by using the method suggested above. If this is done then $B = 0.1155$.

Using $B = 0.1155$ values for $f_X(x)$ can be obtained for a suitable range of X values and are shown plotted in Fig.5.3.

Example 5.5

Assuming that the mean and standard deviation values of example 5.1 are values of a saturated clay's unit cohesion, c_u, and that the minimum and maximum values of c_u are 22 and $55 kN/m^2$ respectively, fit the results to a beta distribution.

Solution

$a = 22$ and $b = 55$, therefore the distribution is asymmetrical with the lower tail longer than the upper, i.e. it has a negative skew.

$$m_X = 44.85\text{kN/m}^2 ; \quad \sigma_X = 6.56\text{kN/m}^2$$

$$\Rightarrow \quad 44.85 = 22 + \frac{r}{t}(55 - 22) = 22 + 33\frac{r}{t} \quad \ldots\ldots\ldots\ldots\ldots(A)$$

$$\sigma_X^2 = 6.56^2 = (55 - 22)^2 \frac{r(t - r)}{t2(r + 1)}$$

Hence t = 4.39 and, from(A), r = 3.039

$$\text{For } r = 3 \text{ and } t = 4, \; B = \frac{(3-1)!(4-3-1)}{(4-1)!} = 0.3333$$

$$\text{For } r = 3 \text{ and } t = 5, \; B = \frac{(3-1)!(5-3-1)!}{(5-1)!} = 0.0833$$

Approximate value for B = 0.3333 - 0.39(0.3333-0.0833) = 0.2358

Using the suggested method gives B = 0.1850 and, with this value, the pdf is shown in Fig.5.4.

Soil distributions

Note - the following discussion applies to both compacted and to undisturbed soils.

Various workers have been involved in the study of the forms of distribution that different soil parameters take together with the determination of the average value that their coefficients of variation take. The main workers in this field have been Lumb (1966 and 1970), Schultze (1972), Turnbull et al. (1966), Meyerhof (1970), Hooper & Butler (1966) and Alonso (1976).

At the present time the consensus of opinion is broadly as set out below.

Many soil parameters have distributions that approximate to the normal, or log normal, distribution. These parameters include:-

Density (dry and bulk)
Voids ratio
Water content
Degree of saturation
Liquid limit
Plastic limit
Plasticity index
Particle specific gravity
Coefficient of consolidation

Parameters whose distributions do not often approximate very well to normal distributions include:-

Compression Index
Coefficient of volume decrease
Permeability
Angle of friction
Cohesion

These notes will only consider soil strength and the reader is referred to the listed references for information on the other parameters.

Distributions of soil strength parameters

1 - Two component soils

The strength of unsaturated soils, such as silts, sandy clays, etc., are made up from two components :-

i) - the unit cohesion, c
ii) - the tangent of the angle of friction, $\tan\phi$

For such soils the drained strength parameters give the best measure of strength and it is should be noted that the symbols c and ϕ used here refer to the drained, or effective, cohesion and angle of friction of the soil.

i) - Cohesion - this parameter best fits a positively skewed beta distribution. (Positively skewed means that the upper tail of the distribution is longer than the lower tail).

It is apparent from the formula for $f_X(x)$ for the beta distribution that the minimum and maximum values, a and b, of the distribution must be known. Provided that these values have been obtained by test measurements there will be no problem but, bearing in mind that there will need to be about 30 measurements in order to obtain values of a and b to an acceptable level of significance, some form of estimation will generally be necessary.

For the asymmetrical beta distribution of the cohesion of a two component soil Lumb, (1970) suggests the following approximate values for a and b:-

$$a = m_c - 1.9\sigma_c; \quad b = m_c + 4.1\sigma_c$$

ii) - Friction - the distribution of the tangent of the effective angle of friction is more or less symmetrical and is close to a symmetrical beta distribution.

If no measured values of the minimum and maximum values of $\tan\phi$ are available Lumb suggests the following approximations:-

$$a = m_{\tan\phi} - 2.3\sigma_{\tan\phi}; \quad b = m_{\tan\phi} + 2.3\sigma_{\tan\phi}$$

2 - One component soils

i) - Cohesion

For a saturated clay the relevant strength component is the undrained cohesion, c_u, which is completely different to the drained cohesion and, as one would intuitively expect, it has an entirely different distribution.

The best approximation to this distribution is a negatively skewed beta distribution, i.e an asymmetrical distribution with a longer tail of lower values.

Approximate values for the minimum and maximum values are:-

$$a = m_{cu} - 3.5\sigma_{cu}; \quad b = m_{cu} + 1.6\sigma_{cu}$$

ii) - Friction

It has been found that the assumption of a symmetrical beta distribution ,as for a two component soil, or of a normal distribution both give satisfactory approximations for the distribution of tanϕ, where ϕ is the drained, or effective angle of shearing resistance, although the beta distribution gives a closer fit for the lower tail.

Note

It has been shown, Lumb, (1965) that the effective angle of friction, ϕ, is statistically independent of grading parameters, voids ratio and degree of saturation and that both ϕ or tanϕ can be taken as being normally distributed variables.

Percentage coefficients of variation of some soil variables

	Sand	Silt	Clay
Density			
(undisturbed soil)	5 - 10	5 - 10	5 - 10
(compacted soil)	2.5 - 5	2.5 - 5	2.5 - 5
Water content			
(undisturbed soil)	5	10 - 23	12 - 22
(compacted soil)	5	5 - 12	6 - 11
Void ratio	13 - 30	22	15 - 32
Liquid limit		5.5	22 - 28
Plastic limit		12	20 - 45

Strength - one component soil

Soft clay (cu < 40kN/m^2) - undrained cohesion ... V = 20 - 25%

Hard clay (cu > 40kN/m^2) - undrained cohesion .....V = 20 - 35%

Sand - drained angle of shearing resistance ....V = 5 - 15%

Strength - two component soil

Two component soils are extremely variable in strength. Typical coefficients of variation are given by Lumb, (1974):-

Clay shale

cohesion 95%; $\tan\phi$ = 46%

Cohesive till

cohesion 100%; $\tan\phi$ = 18%

Residual sands and silts

cohesion 17%; $\tan\phi$ = 6%

CHAPTER SIX - LOADINGS

The soil parameters considered so far have been single variables with values assumed to be independent of time. This assumption can usually be justified in the reliability analyses of geotechnicial structures.

For example, although the shear strength of a soil changes from undrained to drained, a reliability analysis is usually carried out using either the undrained or the drained strength values, depending upon the design requirements.

The strengths and densities of constructional materials are assumed to be non-time dependent, any variations being considered as a random process of the type discussed in the previous chapter.

The dimensions of a structure, for all practical purposes, do not vary with time and are often assumed constant in reliability work.

However one must not ignore the possibility of finishing layers being progressively placed on top of existing layers leading to the phenomenon of a gradually increasing dead load.

Superimposed loadings are obviously time dependent and can vary considerably in their nature although there can be cases of predictable floor loadings where the assumption of a non-time dependent normal or lognormal distribution is realistic and will lead to satisfactory results.

Outside the above exception the effect of time on superimposed loading should be considered.

A superimposed load may involve a sudden change followed by a period in which the magnitude of the loading barely alters, such as the change in furniture and equipment loadings caused by the requirements of a new tenant in a rented industrial or office building.

Such loadings have been discussed by Tang, (1981), and their consideration is necessary in both the evaluation of long term effects such as settlement and in the study of how the structure is likely to behave when subjected to some form of extreme loading.

Extreme values of loading can be created within the continual and irregular series of load values set up by a natural agency, such as snow, wind waves and earthquakes.

For design the main concern is about the largest, or extreme, load values that the structure is likely to be subjected to together with the minumium resistance that the structure will offer. This aspect of reliability analysis is considered in the next section.

Theory of extreme values

For such loadings we are obliged to consider the statistical theory of extreme values.

As already mentioned extreme load values can arise from several different and independent causes, probably the most important being earthquakes, wind, waves and snow.

The formation of an extreme value distribution can be illustrated numerically, as in the following example which deals with the possible variation of the value of wind velocity.

Example 6.1

At a certain geographical location the maximum weekly wind speed was noted for a period of a year and was found to be normally distributed with a mean of 65km/hr and standard deviation of 16km/hr.

Prepare a histogram of the maximum annual wind speed for a period of 50 years.

Solution

If, of course, there was a 50 year set of weekly readings then the maximum value for each year could be obtained and a histogram plotted.

Not having such readings, but knowing that the weekly values are normally distributed, an approximation to the histogram can be obtained by Monte Carlo simulation.

The minimum and maximum wind speeds recorded will be assumed to be the mean value ± 4 x the standard deviation, i.e. 1 and 127km/hr respectively.

Hence, if the computer is programmed to produce 10,000 random numbers between 1 and 10,000 then the probability of generating the actual number 1 is 0.0001 which, from statistical tables of the cdf of the standard normal distribution, gives $z = 3.875$ and is therefore the probability of experiencing a wind speed of $65 - 3.875 \times 16 = 3$km/hr.

Similarly the probability of producing the number 10,000 is also 0.0001 and represents the probability of experiencing a wind speed of 127km/hr.

Within these two extremes a whole range of probable wind speeds can be generated. If, for example the number generated lies beween 1056 and 1587 then it lies between the probabilities of 0.1056 and 0.1587, corresponding to z values of -1.25 and -1.00. Hence, a number generated within this range is taken to represent a wind speed of $65 - 1.125 \times 16 = 47$km/hr.

The programme was prepared to produce 50 independent distributions of 52 independent values, to represent the 52 weeks of each of the 50 years. The programme could easily have been adusted to produce 50 sets of 365 values, to represent daily readings but it was felt that such a large number of values would only confuse the example.

A complete set of weekly readings for one year are set out below and are illustrated in Fig.6.1.

Week	Wind speed (km/hr)	Week	Wind speed (km/hr)	Week	Wind speed (km/hr)
1	63.0	2	63.0	3	71.0
4	39.0	5	59.0	6	75.0
7	71.0	8	79.0	9	71.0
10	87.0	11	63.0	12	111.0
13	75.0	14	119.0	15	75.0
16	75.0	17	87.0	18	51.0
19	51.0	20	71.0	21	111.0
22	15.0	23	71.0	24	87.0
25	71.0	26	55.0	27	59.0
28	67.0	29	67.0	30	127.0
31	43.0	32	111.0	33	67.0
34	67.0	35	75.0	36	39.0
37	87.0	38	63.0	39	67.0
40	51.0	41	51.0	42	71.0
43	59.0	44	67.0	45	59.0
46	67.0	47	71.0	48	55.0
49	51.0	50	43.0	51	51.0
52	79.0				

It is seen that, for this year, the maximum wind velocity was 127.0km/hr.

The histogram is shown in Fig.6.2 and demonstrates that the wind speed distribution is normal.

The procedure was now repeated another 49 times to yield the 50 values of maximum annual wind speed. Each of these values is a measurement of a new random varible Y, the variable that represents the extreme value of wind speed.

These values are tabulated below.

Year	Max. speed (km/hr)	Year	Max. speed (km/hr)	Year	Max. speed (km/hr)
1	87.0	2	103.0	3	115.0
4	95.0	5	95.0	6	99.0
7	111.0	8	99.0	9	103.0
10	99.0	11	99.0	12	99.0
13	103.0	14	91.0	15	91.0
16	95.0	17	103.0	18	91.0
19	111.0	20	107.0	21	107.0
22	99.0	23	127.0	24	103.0
25	107.0	26	91.0	27	111.0
28	107.0	29	103.0	30	99.0
31	99.0	32	91.0	33	99.0
34	107.0	35	95.0	36	111.0
37	99.0	38	103.0	39	119.0
40	99.0	41	103.0	42	127.0
43	95.0	44	99.0	45	107.0
46	103.0	47	107.0	48	91.0
49	99.0	50	111.0		

The histogram of these values of Y is shown in Fig.6.3A. From this histogram it is possible to obtain an approximation to the pdf of this particular distribution and this is drawn in on the figure.

It is seen that although the parent distributions were all normal this distribution of extreme values is by no means normal. It is referred to as an extreme distribution.

By considering the pdf line drawn on Fig.6.3A it is possible to slightly adjust the arrangement of the histogram cells so that they conform more closely with the sketched in pdf line. This has been done in Fig.6.3B and it is now a simple matter to obtain an approximation for the $f_Y(y)$ and $F_Y(y)$ values bearing in mind that the total area of the histogram is equal to 1.0.

There are 50 segments in the 11 cells. The first cell consists of one segment and the area of this cell is therefore 1/50 = 0.02.

Dividing the area of the cell by its width (4km/hr) gives the central height of the cell, i.e. $f_Y(y)$ and is therefore 0.005.

$F_Y(y)$ is obviously equal to 0.02.

The second cell has 3 segments so that $f_Y(y) = 3/50x4 = 0.015$ and $F_Y(y) = (3+1)/50 = 0.08$.

The $f_Y(y)$ and $F_Y(y)$ values are set out below.

y (km/hr)	$f_Y(y)$	$F_Y(y)$
87	0.005	0.02
91	0.015	0.08
95	0.035	0.22
99	0.06	0.46
103	0.05	0.66
107	0.03	0.78
111	0.02	0.86
115	0.015	0.92
119	0.01	0.96
123	0.005	0.98
127	0.005	1.00

Extreme value distributions

Extreme distributions can obviously be obtained for either tail of the parent distribution, i.e. for maximum or minimum values.

Probably the most valuable work carried out on extreme value statistics is that by Gumbel, (1958).

The distribution from which the extreme values are generated is known as the parent distribution, an example being the yearly set of the weekly wind values of the preceeding example.

The distribution of extremes can be easily expressed as a function of the parent distribution and the sample size, n. (i.e the number of times that the distribution was considered = 50 in example 6.1)

The probability that all of the n independent obsevations will be less than x is $P^n(x)$. This can be re-expressed as the probability that y, the largest value among the n independent observations, is less or equal to x, $\Phi_n(y)$.

Hence $\Phi_n(y) = P^n(x)$

In other words the cumulative distribution function of the extreme value distribution is the cumulative distribution function of the parent distribution raised to the power of the sample size.

Unfortunately the formula is of little practical significance owing to the high powers generally involved.

The significant property of the extreme distribution was discovered by Fisher and Tippet, (1928), who showed that, as the sample size increases the actual form of the distribution asymptotically approaches one of three distinct forms, called Type I, Type II and Type III.

Fisher and Tippet showed that the type of distribution that is finally achieved depends upon the properities of the tails of the parent distribution.

The Type I distribution

If the parent distribution consists of a random variable X with a cumulative probability distribution of the form:-

$$F_X(x) = 1 = \exp[-g(x)]$$

where g(x) is a monotonic increasing function of x (see Chapter 5), then the extreme value distribution of Y is Type I.

(The normal and beta distributions have cdfs of this form.)

When considering the pdf of the Type I distribution it is advantageous to think in terms of a reduced variable, $\alpha(y - u)$.

u is the mode of the distribution and α is a measure of the dispersion.

It can be shown, Benjamin & Cornell, (1970), that:-

$$\alpha = \frac{1.282}{\sigma_Y}; \qquad u = m_Y - \frac{0.577}{\alpha}$$

The expressions for $f_Y(y)$ and $F_Y(y)$ then become:-

$$f_Y(y) = \exp[-\alpha(y-u) - \exp(-\alpha(y-u))]$$

$$F_Y(y) = \exp[-\exp(-\alpha(y-u))]$$

For maximum values the distribution is positively skewed and has the form of the pdf estimated in Fig. 6.3B. For minimum values the distribution is negatively skewed and can allow for values of y being less than 0.

Note

It is generally agreed that the Type I extreme distribution is the most suitable for civil engineering design work, which is usually concerned with the prediction of maximum values of snow and wind loading. It has also been used to model the long term distribution of North Sea wave heights, Saetre, (1975)

However, for the sake of completeness, brief notes on the Type II and the Type III distributions are set out below.

Type II distribution

If the parent distribution has a cdf of the form:-

$$F_X(x) = 1 - C\left(\frac{1}{x}\right)^k \qquad x \geq 0$$

then the extreme distribution is Type II with the following expressions for $f_Y(y)$ and $F_Y(y)$:-

$$f_Y(y) = \frac{k}{u}\left(\frac{u}{y}\right)^{k+1}\exp[-(u/y)^k] \qquad y \geq 0$$

$$F_Y(y) = \exp[-(u/y)^k] \qquad y \geq 0$$

A relationship exists between the Type I and Type II distributions which is identical to the relationship between the normal and the lognormal distributions.

It can be shown that if Y has a Type II extreme distribution then the variable $Z = \ln(Y)$ has a Type I extreme distribution.

Type III distribution

If the parent distribution has a fixed limit on its maximum value, $x_{max} = w$ (say), so that the cdf of the parent distribution, in the region of w, has the form:-

$$F_X(x) = 1 - c(w - x)^k \qquad x \leq w, \quad k > 0$$

then the extreme distribution is Type III with the following expressions for $f_Y(y)$ and $F_Y(y)$:-

$$f_Y(y) = \frac{k}{w - u}\left(\frac{w - y}{w - u}\right)^{k-1} \exp\left[-\left(\frac{w - y}{w - u}\right)^k\right] \qquad y \leq w$$

$$F_Y(y) = \exp\left[-\frac{w - y}{w - u}^{k}\right] \qquad y \leq w$$

Note

Because of the different types of earthquake sources, i.e. fault lines, aerial sources and point sources, the maximum lateral force to which an earth dam may be subjected because of earthquake action cannot be modelled satisfactorily by the above extreme distributions.

Analytical models whereby the maximum acceleration at a site can be estimated have been prepared by Cornell, (1968) and DerKiureghian & Ang, (1977) but a study of this aspect of reliability analysis is outwith the scope of this thesis.

Example 6.2

Measurements taken over several years at a geographical location indicate that the mean value and standard deviation of the maximum annual wind velocity are 102.3 and 8.5km/hr. repectively.

Assuming a Type I extreme distribution, plot the pdf and the cdf of the wind speed's maximum values.

Note

The values of the mean and standard deviation are those of the set of simulated maximum values obtained in example 6.1.

Solution

$$\alpha = \frac{1.282}{\sigma_Y} = \frac{1.282}{8.5} = 0.1508$$

$$u = m_Y - 0.577/\alpha = 102.3 - 0.577/.1588 = 98.66\text{km/hr}$$

Substituting suitable values for Y into the expressions for $f_Y(y)$ and $F_Y(y)$ leads to the following results:-

y	$f_Y(y)$	$F_Y(y)$
87	0.004	0.004
91	0.022	0.046
95	0.047	0.185
99	0.055	0.397
103	0.046	0.603
107	0.032	0.759
111	0.020	0.860
115	0.012	0.921
119	0.007	0.956
123	0.004	0.976
127	0.002	0.987

The cdf and pdf plots are shown in Fig.6.4 along with the estimated plots obtained in example 6.1.

Soil induced loading

A major portion of the loading carried by a soil structure is generated by the soil itself.

Examples of these induced forces that immediately spring to mind are:-

Vertical forces - due to the weight of the soil -affect sliding frictional resistance, settlement, negative skin friction on piles, etc.

Lateral forces - thrusts due to active or passive earth pressure.

It appears to the writer that all of these induced soil forces are automatically allowed for in the reliability approach he has suggested. (See, for instance, examples 4.1 and 4.2).

However there is one major soil variable that has so far not been mentioned and that is pore water pressure.

In its simplest form the problem of pore water pressure is one of statics and concerns the hydrostatic state of soil below a ground water level. The pore water pressure in the soil is in a state of equilibrium and varies as the ground water level varies.

Ground water level variations are due to several agencies, such as seasonal temperature variations, rain and snow falls, constructional, irrigation and pumping schemes, etc.

If records, such as piezometric readings, have been kept over a period it is possible to determine the maximum and minimum levels of ground water and to then fit them to some form of extreme distribution.

Such a situation is fairly unlikely and, for large and important structures, hydrological models must be used to predict ground water level changes, Uno et al, (1981).

For relatively small structures a ground water level is usually obtained from the site investigation report. Generally an estimation as to the highest ground water level can be made and this value is then taken as a constant in the reliability analysis.

For slopes the coefficient of variation of the pore water pressure,and hence of r_u, is typically some 5 to 10%, Yuceman & Tang, (1975).

Seepage of water through soil leads to seepage forces related to both the hydraulic gradient and the permeability of the soil. It is generally accepted that any measured values of permeability can only be considered to be within plus or minus one order of magnitude and this can have a large effect on the accuracy of flow nets.

Mostyn, (1983) has suggested that the use of the geometric, rather than arithmetic, mean of the permeability measurements can lead to more realistic flow diagrams.

However when one considers how the inhomogenous nature of soil deposits, particularly silts and soils with erratically positioned sand lenses, can lead to further uncertainties it is seen that the prediction of seepage of seepage forces is an extremely difficult problem and one on which much research work remains to be done.

CHAPTER SEVEN - TREATMENT OF VARIABLES WITH NON-NORMAL DISTRIBUTIONS

If it is known that some of the variables involved in a reliability analysis have distributions that are not normal then the accuracy of the analysis can be increased if these variations are allowed for.

Fiessler and Rackwitz, (1976), proposed a method for the treatment of non-normal random variables and it is this method that is used in these notes.

In chapter 2 it was explained that, in the second moment method of reliablity analysis, the limit state function, Z, is approximated lineally at the design point, x^{*}.

Fiessler and Rackwitz showed that the non-normal distribution of a random variable, X, can be approximated at the design point to an equivalent normal distribution that has the same cumulative probability value at the design point as the orginal variable.

If the standardised variable of this equivalent distribution that corresponds to X is given the symbol E then:-

$$E = \frac{x - m_X^N}{\sigma_X^N}$$

m_X^N and σ_X^N are respectively the mean and the standard deviation of the equivalent distribution.

$$\text{i.e.} \quad m_X^N = x - E\sigma_X^N$$

Fiessler and Rackwitz showed that $\sigma_X^N = f_X(x)/f_E(\varepsilon)$

Hence the expressions for σ_X^N and m_X^N are:-

$$\sigma_X^N = \frac{\int^N [\Phi^{-1}(FX(x^*))]}{f_X(x)}$$

$$m_X^N = x - \Phi^{-1}(F_X(x^*))$$

where $F_X(x^*)$ = the cumulative probability of X at x^*

$f_X(x^*)$ = the probability density of X at x^*.

$\Phi^{-1}(.)$ = the inverse normal distribution

$\int^N(.)$ = the standardised normal density function

Values for the last two items can be obtained from published tables (or a suitable subroutine).

The values of m_X^N and σ_X^N are used instead of m_X and σ_X in the iteration procedure for determining β.

Distributions of related variables

Many of the basic variables that make up a geotechnical limit state function are related to the angle of shearing resistance, ϕ.

Variables such as K_a, N_c, N_q, N_γ, i and e are all functions of ϕ but this does not necessarily mean that the forms of their distributions will be the same as that of ϕ.

If required the form of the distribution of a function of ϕ can be obtained by means of the Monte Carlo method. This technique has already been described in connection with the determination of failure probabilities, see chapter 2.

To determine the pdf of such a variable the probabilities of occurrence of values of ϕ are first obtained, from the cumulative probability function of ϕ, $F_\phi(\phi)$,

These probabilities are next related to a set of generated random numbers so that a set of random values of ϕ can be obtained. For each ϕ value generated a corresponding value of the related variable can be calculated and, provided a sufficient number of iterations are carried out, a histogram which will give a close approximation to the pdf of the variable can be drawn.

The pdf of N_γ , as used in example 3.7, was obtained by means of the Monte Carlo technique and is shown in Fig.7.1. The pdfs of N_c and N_q have a similar form.

The mean and standard deviations of these three coefficients obtained from the Monte Carlo method , using 2000 iterations each, were:-

Coefficient	Mean	s.d.
N_c	31.35	8.10
N_q	19.71	7.01
N_γ	17.09	8.47

Example 7.1

In example 3.7 the three bearing capacity coefficients, N_c, N_q and N_γ were all assumed to have normal distributions.

At this stage it is of interest to attempt to determine the form of the actual distributions of these variables and to examine the effect on the value of β when any departures from normality are allowed for.

Solution

As established in chapter 5 the probability distribution of the angle of shearing resistance is almost equally well represented by the assumption of a normal distribution or by the assumption of a symmetrical beta distribution.

The assumption of a beta distribution for ϕ leads to a more meaningful tail representation and will be assumed here.

The Monte Carlo simulation of the pdf of N_γ, Fig.7.1, has already been stated to have a form very similar to that of the pdfs for Nc and Nq.

The mean and standard deviation values of the bearing capacity coefficients, 30° and 3°, have already been established and it is simple matter to fit them to a lognormal or to an extreme Type I distribution, using the methods outlined in chapter 5.

In order to fit the mean and s.d. values to a beta distribuiton the maximum and minimum values of the various coefficients must be known.

Using Lumb's suggestions for the a and b values for a symmetrical beta distribution of ϕ, the drained angle of shearing resistance, gives minimum and maximum values for ϕ of 23° and 37°.

The maximum and minimum values of the bearing capacity coefficients can quickly be found from Appendices I, II and III:-

Coefficient	Mean	s.d.	Minimum	Maximum
N_c	30.14	7.35	18.05	55.63
N_q	18.4	6.45	8.66	42.92
N_γ	15.07	7.56	4.88	47.38

The three fitted distributions are shown in Fig.7.2 and it is fairly obvious that the pdfs of the three bearing capacity coefficients are close to both the lognormal and the extreme Type I distributions.

The writer considers that, as we are dealing with soil parameters and not loads, the assumption of lognormal distributions is more acceptable.

Treatment of N_c

For the mean value of X (30.14):-

$$f_X(x) = 0.0547 \text{ and } F_X(x) = 0.5478$$

$$\Phi^{-1} F_X(0.5478) = 0.1189; \quad \int^{N} [\Phi^{-1} F_X(0.5478)] = 0.3961$$

$$\sigma_{Nc}^{N} = \frac{0.3961}{0.0547} = 7.24; \quad m_{Nc}^{N} = 30.14 - .1189 \times 7.24 = 29.28$$

Treatment of N_q

For the mean value of X (18.40):-

$$f_X(x) = 0.0628 \text{ and } F_X(x) = 0.5675$$

$$\Phi^{-1} F_X(0.5675) = 0.1684; \quad {}^{N}[\underline{O}^{-1} F_X(0.5675)] = 0.3933$$

$$\sigma_{Nq}^{N} = \frac{0.3933}{0.0628} = 6.26; \quad m_{Nq}^{N} = 18.4 + 0.1684 \times 6.26 = 17.35$$

Treatment of N_γ

For the mean value of X (15.07):-

$$f_X(x) = 0.0729 \text{ and } F_X(x) = 0.5749$$

$$\Phi^{-1} F_X(0.5749) = 0.1872; \quad \int^{N} [\Phi^{-1} F_X(0.5749)] = 0.3920$$

$$\sigma_{N\gamma}^{N} = \frac{0.3920}{0.0729} = 5.38; \quad m_{N\gamma}^{N} = 15.07 - 0.1872 \times 5.38 = 14.06$$

Hence the means and s.d values of the variables have become:-

Variable	Symbol	Mean	s.d.	units
γ	X_1	19	0.5	kN/m^3
c	X_2	90	30	kN/m^2
N_c	X_3	29.28	7.24	
N_q	X_4	17.35	6.26	
N_γ	X_5	14.06	5.38	
P	X_6	25,000	2500	kN

With these values the iteration procedure for β gives:-

Iteration	y_1	y_2	y_3	y_4	y_5	y_6	β	g(y)
1	0	0	0	0	0	0	0	45606
2	-0.08	-1.65	-1.22	-0.86	-0.21	0.28	2.25	7395
3	-0.10	-2.03	-0.97	-1.51	-0.37	0.49	2.78	-1433
4	-0.07	-2.09	-0.66	-1.43	-0.35	0.46	2.68	-333
5	-0.07	-2.15	-0.58	-1.34	-0.33	0.43	2.66	-74.9
6	-0.07	-2.17	-0.53	-1.32	-0.33	0.43	2.65	-12.8
7	-0.07	-2.18	-0.52	-1.31	-0.32	0.42	2.65	-2.4
8	-0.07	-2.18	-0.51	-1.30	-0.32	0.42	2.65	-0.45

The reliability index (assuming that the bearing capacity coefficients have lognormal distributions) = 2.65.

It is seen that this value is extremely close to that obtained from full correlation, 2.60 (see chapter 3) and indicates that the proposed method becomes even more dependable when departures from normality are allowed for.

As a matter of interest when the bearing coefficients are fitted to an extreme Type I distribution the value for the reliability index reduces to 2.33 which is an indication that the coefficients follow a distribution that is closerto lognormal.

Example 7.2

Determine the value for the reliability index of example 3.7 assuming that the variables have the following distributions.

Unit weight of soil	- normal
Angle of friction of soil	- symmetrical beta
Cohesion of soil	- asymmetrical beta
Column load	- extreme type I

Solution

Treatment of cohesion, c

This variable has an asymmetrical beta distribution.

The cohesion is part of a two component soil and Lumb's suggestion, (1970) for its minimum and maximum values will be used.

$$a = m_c - 1.9\sigma_c = 90 - 1.9 \times 30 = 33\text{kN/m}^2$$

$$b = m_c + 4.1\sigma_c = 90 + 4.1 \times 30 = 213\text{kN/m}^2$$

Using the procedure described in chapter 3:-

$$r = 2.150; \quad t = 6.790; \quad B = 0.0312$$

and, for a mean value of X (90kN/m^2)

$$f_X(x) = 0.0119 \quad \text{and} \quad F_X(x) = 0.5391$$

From tables, etc.:-

$$\Phi^{-1} F_X(0.5391) = 0.0982 \quad \text{and} \quad f^N[\Phi^{-1} F_X(0.5391)] = 0.3971$$

Note $f^N[\Phi^{-1} F_X(0.5391)]$ is simply the ordinate of the standardised normal density function at $Z = 0.0982$.

Hence, according to Fiessler and Rackwitz:-

$$\sigma_c^N = \frac{0.3971}{0.0119} = 33.37 \quad \text{and} \quad m_c^N = 90 - 0.0982 \times 33.37 = 86.72\text{kN/m}^2$$

Treatment of column load, P

This variable has an extreme type I distribution.

$$\alpha = \frac{1.282}{2500} = 0.0005128; \quad u = 25000 - \frac{0.577}{0.0005128} = 23874.8$$

Hence, for the mean value of X (25,000kN) and by substitution in the formulae quoted in chapter 5:-

$$f_X(x) = 0.000164 \text{ and } F_X(x) = 0.5703$$

$$\Phi^{-1} F_X(0.5703) = 0.1755 \text{ and } f^N[\Phi^{-1} F_X(0.5703)] = 0.3928$$

$$\text{Hence } \sigma_P^N = \frac{0.3928}{0.000164} = 2395 \text{ and } m_P^N = 25000 - 0.1755 \times 2395$$

$$= 24,580\text{kN}$$

The three bearing capacity coefficients have already been discussed in example 7.1. As in example 7.1 the means and standard deviations of these variables have been transformed on the assumption that they follow lognormal distributions.

Hence the means and s.d values of the variables have become:-

Variable	Symbol	Mean	s.d.	units
γ	X_1	19	0.5	kN/m^3
c	X_2	86.72	33.37	kN/m^2
N_c	X_3	29.28	7.24	
N_q	X_4	17.35	6.26	
N_γ	X_5	14.06	5.38	
P	X_6	24,580	2395	kN

With these values the iteration procedure for β gives:-

Iteration	y_1	y_2	y_3	y_4	y_5	y_6	β	g(y)
1	0	0	0	0	0	0	0	44407
2	-0.07	-1.62	-1.04	-0.76	-0.19	0.24	2.09	6849
3	-0.09	-2.03	-0.66	-1.27	-0.32	0.40	2.54	-1892
4	-0.06	-2.05	-0.35	-1.15	-0.28	0.36	2.42	-386
5	-0.06	-2.08	-0.31	-1.06	-0.26	0.33	2.40	-39.6
6	-0.06	-2.09	-0.29	-1.06	-0.26	0.33	2.40	-2.2
7	-0.06	-2.09	-0.28	-1.05	-0.26	0.33	2.39	-0.2
8	-0.06	-2.09	-0.28	-1.05	-0.26	0.33	2.39	-0.0

The reliability index = 2.39, as opposed to 2.65 in example 7.1.

Hence the probability of failure increases slightly, from 4.0×10^{-3} to 8.4×10^{-3}, when the forms of the distributions are allowed for.

CHAPTER EIGHT - STATISTICAL ASPECTS OF SITE INVESTIGATIONS

Soil and rock deposits, with the properties given to them by nature, are amongst the most variable of the materials that a civil engineer is called upon to use. Faults can occur quite erratically and tests and measurements carried out are of necessity limited, with the often added complication that parameter values from laboratory tests can only approximate the in-situ values.

The standard approach to geotechnical uncertainty is the observational method, evolved by Terzaghi and Peck, (1948) and later fully discussed by Peck, (1969) and Casagrande, (1965).

Briefly the technique consists in designing the foundation, retaining wall, etc. using the information available and then, during construction, check on the original design assumptions by measuring predetermined parameters such as pore water pressures, deformations, etc. In the light of these observations further construction, i.e. rate or dimensions, can be modified.

In the hands of the competent and experienced engineer the observational method will yield satisfactory and economical structures. The method may not be so effective with a relatively inexperienced engineer and, for this reason, factors of safety based on experience gained from similar structures are included in the design.

Statistics and probability theory provide mathematical approaches that can help to deal with uncertainties and be of assistance with the observational method.

Obviously there is nothing truly "random" about the variation of soil properties from one point to another. For example the formation of an estuarine deposit will have been controlled by such agencies as the rate of erosion of the environmental soil, the depth of suspension, the current velocity, etc. although the effect of wind, tidal currents, etc. may well have imposed a random pattern of variation across the overall trend.

Site investigation procedure.

The object of a site investigation or a proposed structure is to determine those properties of the soil that will significantly affect the design and construction. From a statistical point of view this involves obtaining estimates of the means, standard deviations and probability distributions of all relevant parameters.

In order to do this the first part of a site investigation must be to determine whether different soil layers lie beneath the ground surface and, if so, to determine the positions of the various soil horizons. Once the subsoil had been divided into sub-regions then each layer can be regarded as an independent population from which appropriate samples will be collected.

As was confirmed at a recent British Geotechnical Society's debate, Butcher, (1984), except for a few enlightened cases, the average site investigation is presently carried out in blind accordance with a set of arbitrary specification clauses, more with an eye on costs than on the need to obtain relevant information.

It is acknowledged that there have been many cases in the past where poor site investigations have yielded inadequate subsoil information and hence led to costly over-designed structures.

If statistical methods are to be employed in geotechnical design and decision making then the interested parties to a site survey must realise that there will be a need for an increase in the number of soil samples collected if dependable values for means and variances of the relevant soil parameters are to be obtained.

The situation is simply that the greater the number of samples collected and tested then the greater the dependability of the reliability analysis prediction, provided that the samples collected are truly representative of the different sub-regions.

It has been illustrated by Morse, (1971) that classifying soils into different sub-regions by visual inspection and field tests may not be sufficiently accurate and can lead to quite wrong design parameters if different, although superficially similar, soils are grouped as belonging to the one population.

For this reason a site investigation for an important structure should be split into two phases.

In the first phase enough boreholes should be put down and enough soil samples collected so that borehole journals can be prepared and, with the help of laboratory classification tests on the soil samples taken, the substrata modelled as a set of sub-regions.

It may be that in some particular investigation enough information is available from the first phase to complete this work. If this is not so then the second phase would involve a return to the site for further borings and soil sampling to obtain the missing information.

The determination of the various horizons separating the soil deposit into subregions may often prove difficult and can require considerable experience on the part of the engineer attempting to model the subsoil in this manner.

This modelling procedure will have little to do with probability theory and, without the right amount of information, the final estimation of the soil profile, although based on accurate information, could be quite wrong. Fig.8.1A shows the information obtained from two boreholes at some distance apart at a particular site.

The information from these two boreholes is very similar and yet Figs. 8.1B,C and D illustrate three possible interpretations. Statistics can be of little help here. What is required is extra information which can only be obtained from another borehole.

The probability of the existence of a geological feature cannot be predicted by a simple statistical procedure. Geological forms can only be predicted by some form of inductive reasoning possibly enhanced by the use of Baye´s theorem involving any results obtained from the site investigation together with past experience. (See example 1.13).

If it is known or suspected that an old slip surface exists and if it is required that its existence, or non existence, be proved completely, then its location can only be discovered by continuous coring as part of a systmatic search programme.

Soil sampling theory

For the test results obtained from a soil sample to be meaningful the sample must have been taken from the correct soil sub-region.

There is little arguement that for quality control, such as the measurement of density along a compacted embankment or the testing of sand from lorries delivering to a site, sample selection should be done on a random basis.

However it has been shown by Lumb, (1974), that random selection does not give the best results for in-situ sampling. A systematic pattern of boreholes and the collection of samples at equally spaced vertical intervals for each sub-region is the most efficient method.

It is important to realise that in geotechnics the term "sample" means a single specimen of soil or rock whereas in statistics the term "sample" means a set of results or values.

It can generally be assumed that a set of n soil samples taken from a soil layer will yield one statistical value consisting of n measurements of a particular soil parameter obtained from the n soil samples.

Point estimation

In geotechnics we rarely deal with more than one statistical sample per soil stratum which usually consists of 4, 5 or 6 values.

We are therefore usually obliged to estimate the mean of a population, m_X, from only one sample value, m_S. Such an estimation is called a point estimation.

It is useful to be able to roughly estimate how far away the the sample value is from the actual mean value of the population. This is achieved by a term called the standard error of the mean.

The standard error of the mean

As discussed, the value of m_S in soils is usually determined from a small group of measured value.

Let us assume that there is a large number of measurements and that these values are grouped into sets of 5 by some form of random selection.

If m_S is determined for each group then several different values of m_S will be determined.

The distribution of these m_S values will be found to be distributed about a mean value, m_X, and to have a certain variance value. For all practical purposes m_X is the value of the mean of the population.

If the values had been placed in groups of ten the resulting m_S values would still have been distributed about the same mean value, m_X, but would have had a smaller variance, (See Fig.8.2).

In other words the variances of equal groups of values is related to n, the number of values in each group.

This relationship can be expressed as:-

$$Var_n = \frac{\sigma^2}{n}$$

where Var_n = variance of values when grouped with n values per group

σ^2 = variance of values when each value taken separately

Now the standard deviation is the square root of the variance:

$$\Rightarrow \quad \sigma_n = \frac{\sigma}{\sqrt{n}}$$

The term $\frac{\sigma}{\sqrt{n}}$ is the standard error of the mean.

As one would expect, if the values of a population are normally distributed then the sampling distribution of the means, i.e. the distribution of m_S values, will also be normal.

What is of great interest is that, even if the population is not normally distributed, the sampling distribution of the means approaches a normal distribution as the value of n increases.

Therefore, on the assumption that the m_S values are normally distributed, the properties of the normal curve can be used to estimate the dependability of a point estimation of the mean of a population.

The smaller the value of $\sigma/\sqrt{n}$, i.e. the larger the value of n, the more reliable the estimated value of the poulation mean.

It can be shown that for a normally distributed population, with a standard deviation σ, the following levels of confidence apply:-

i) 68.3% probability that m_X lies within the range $m_S \pm \frac{\sigma}{\sqrt{n}}$

ii) 95.4% probability that m_X lies within the range $m_S \pm \frac{2\sigma}{\sqrt{n}}$

iii) 99.7% probability that m_X lies within the range $m_S \pm \frac{3\sigma}{\sqrt{n}}$

<u>Example 8.1</u>

A sample of values taken from a normally distributed population is set out below.

6.8, 7.2, 5.4, 8.9, 10.2

If the standard deviation of the population is 2.1 determine the range of the sample values within which the mean of the population has a 95.4% probability of lying.

Solution

$m_S = 7.7$ Hence range of sample values within which m_X has a 95.4% chance of occurring =

$$7.7 \pm 2 \times \frac{2.1}{\sqrt{5}} = 5.82 \text{ to } 9.58$$

<u>Note</u>

Obviously other values of the multiplier of $\sigma/\sqrt{n}$ apply for other probability values. An important one is 1.645 which corresponds to the 95% reliability value.

Estimation of the standard deviation of a population

Although a value was given in example 8.1 the standard deviation of a population is not generally known and it is necessary to estimate its value from the test results available.

The estimation often used is known as the Bessel correction and is:-

$$\hat{\sigma} = s\sqrt{\frac{n}{n-1}}$$

Where s = the standard deviation of the n sample values

$$= \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{n} - m_S^2}$$

The use of the circumflex indicates that the value of σ is estimated.

If desired $\hat{\sigma}$ can be calculated directly from the sample values:-

$$\hat{\sigma} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - m_S)^2}{n-1}}$$

It is seen that the effect of Bessel´s correction is to arrive at a value for $\hat{\sigma}$ that is slightly greater than s.

The sample variance, s, tends to underestimate the population variance because in its determination the deviations of the sample values are expressed as the differences from the mean sample value, m_S.

Generally the sum of the squares of the deviations of the sample values about the population mean are slightly greater than if taken about the sample mean and Bessel's correction is important when n is a small number (as in geotechnics).

Student's t distribution

Let the ratio $\dfrac{\text{Error in the sample mean}}{\text{Standard error of the mean}} = z$

$$\text{Then } z = \frac{|^{m}X - {}^{m}S|}{\dfrac{\sigma}{\sqrt{n}}} = \frac{|^{m}X - {}^{m}S|\sqrt{n}}{\sigma}$$

where z = the standardised error between the sample mean and the population mean.

From the ealier discussion it will be appreciated that the distribution of a group of standardised mean errors, taken from a normally distributed population are values of a variable Z which will also be normally distributed.

The above statement is correct only as long as the value used for σ is the true value of the standard deviation of the population.

Generally an estimated value, $\hat{\sigma}$, has to be used in place of σ which means that, instead of establishing a variable Z we establish a variable T where $t = \dfrac{|^{m}X - {}^{m}S|}{\hat{\sigma}/\sqrt{n}}$

$$\text{or} \qquad t = \frac{|^{m}X - {}^{m}S|\sqrt{n-1}}{s}$$

T has a distribution similar to, but not the same as the normal distribution and was discovered by W. S. Gosset who published his findings under the pseudonym A Student.

The value of $\hat{\sigma}$ gets closer to the value of σ as the value of n is increased. For value of n = 30 the t distribution becomes identical to the normal distribution.

In geotechnical work there is little chance of the value of n approaching anything like 30. It is therefore more realistic, when dealing with significance levels, to accept the fact that the probability values of the t distribution should be used in place of values from the normal distribution.

The practical limits for t are from -5 to +5 and the probability corresponding to a particular t value can be found provided that the degrees of freedom value, v, is also known.

The degrees of freedom of a set of values is the number of these values that can be of any magnitude, within the constraint of the calculations about to be carried out on them. For this particular aspect of statistics we can say that $v = (n - 1)$.

Tabulated values of $f_T(t)$ and $F_T(t)$ are readily available and it is therefore a relatively simple matter to obtain the probability value corrresponding to particular t and v values.

Some years ago the writer prepared the graphs reproduced in Fig. 8.3, taken from Smith & Pole, (1980), for the prediction of geotechnical significance levels. With these graphs there is no need to look up tables. Their use is illustrated in examples 8.2 and 8.3.

Example 8.2

i) A sample consisted of the following values:-

10, 8, 6, 4

Determine the range of values within which the true value of the mean of the population has a 95% probability of lying.

ii) In order to increase accuracy a further four values were obtained. These were:-

4.6, 6.8, 8, 9.5

With this extra information determine the range of values within which the mean value has a 95% probability of lying.

Solution

The sample sizes for both (i) and (ii) are small ($n < 30$) and we must therefore use the properties of the t distribution.

i) $m_S = \frac{10+8+6+4}{4} = 7.0$; $s = 2.236$ $\Rightarrow \hat{\sigma} = \sqrt{\frac{4}{3}} \times 2.236 = 2.58$

$v = 4 - 1 = 3$ and, from Fig.8.3, for $P = 95\%$ and $v = 3$, $t = 3.2$

$\Rightarrow$ range of values within which m_X will lie (for $P = 95\%$) is:-

$$7.0 \pm 3.2 \times \frac{2.58}{\sqrt{4}} = 2.87 \text{ to } 11.13$$

ii) $m_S = \frac{4.0+4.6+6.0+6.8+8.0+8.0+9.5+10}{8} = 7.11$; $\hat{\sigma} = 2.170$

$v = 8 - 1 = 7$ and, from Fig.8.3, for $P = 95\%$ and $v = 7$, $t = 2.4$

$\Rightarrow$ range of values within which m_X will lie (for $P = 95\%$) is:-

$$7.11 \pm \frac{2.4 \times 2.170}{\sqrt{8}} = 5.27 \text{ to } 8.95$$

Minimum sample number

From the soil samples obtained during the first phase of a site investigation it is possible to determine whether or not the number of samples collected from each sub-region was sufficient for the required accuracy of prediction. If not then the second phase of the site investigation will be necessary so that further samples can be obtained.

The minimum samples that have to be collected from a sub-region depends upon various factors, not the least being the accuracy of prediction asked for by the design engineer.

If he demands that the average value of the test results should equal the average in-situ value he is demanding the impossible as only an infinite number of samples could satisfy this condition.

Fortunately most engineers will accept a test value that is within 10% of the average in-situ value. This means that the area under the probability curve between the limits set by the engineer must be 0.9. In other words the maximum permissible error at either end of the distribution is 5%, i.e. the significance level is 95%.

Example 8.3

Five undisturbed samples were taken from a stiff clay deposit. Undrained triaxial tests on these samples gave the following values for c_u, the undrained shear strength values:-

100, 80, 95, 110, 100 kN/m^2

Determine the minimum number of samples of the clay that should be taken if the average in-situ undrained shear strength value is to be within 10% of the mean test result.

Solution

With the five samples v = 4. From Fig.8.3 for P = 95%, t = 2.7

Now the mean c_u test result = $97kN/m^2$ and $\hat{\sigma} = 10.95kN/m^2$.

=> Existing range of values for P = 95% = $97 + 2.7 \times 10.95/\sqrt{5}$

= 83.77 to $110.2kN/m^2$ (some 15% either side of the mean)

Obviously more samples are required if the range is to be only 5% on either side of the mean.

Assume, for the moment, that the value of t remains at 2.7

Then m_X must be equal to $m_S \pm 2.7S.E. = m_S(1.0 + 0.1)$

$$\text{i.e. } 2.7\frac{\hat{\sigma}}{\sqrt{n}} = 0.1m_S = 9.7$$

$$\Rightarrow \sqrt{n} = \frac{2.7 \times 10.96}{9.7} = 3.05$$

$$\Rightarrow \quad n = 9.30 \quad (=10)$$

The calculation must now be repeated in an iterative manner.

From Fig.8.3 the value of t that corresponds to v = 9 with a 95% probability is 2.3.

With this value for t the number of samples, n, works out at 6.75. With further iteration it is found that n, the minimum sample number required to satisfy the conditions, = 8 meaning that at least three further samples of the clay must be obtained and tested.

Note

To estimate the variance of the population to the same precision as the mean requires a much larger sample and is not usually attempted in soil and rock reliability analyses.

Probabilistic treatment of the substrata

It has been mentioned that a major part of a site investigation is to divide the subsoil into a set of idealised sub-regions of different soil types.

Only when this procedure has been completed is it possible to carry out meaningful design calculations.

In many cases a straightforward design procedure can be achieved by assuming that these soil layers are locally homogeneous, i.e. they exhibit no significant variations either in thickness or properties within the area of the site.

There are, however, many occasions when the assumption of homogenity is unrealistic. An obvious example is that of a compressible soil layer whose thickness varies randomly.

However, once the various soil horizons have been established, any soil profile characteristic whose uncertainty might have an important effect on the performance predictions can be treated as a random function of either the vertical and/or the horizontal directions, Vanmarcke, (1977).

In theory, given sufficient sample measurements, it would be possible to predict the value of a soil parameter at any location.

However, for economical reasons, the number of test results available will always be limited with the attendant need for some form of interpolation.

Because of this the models of soil deposits that are used in reliability analyses are invariably based on one of two basic assumptions.

i) Soil mass is spatially uniform

A soil mass can be considered to be spatially uniform when it can be assumed that the measured values of a particular parameter vary randomly within it, about the mean value of the measurements.

Throughout the previous chapters spatial uniformity has been tacitly assumed, i.e. the measured values of soil parameters have been assumed to vary randomly with no specific directional trends.

There will be many occasions when the results of a site investigation indicate that the assumption of spatial uniformity will be realistic for design purposes.

There is always the problem, particularly with small structures, when the provisional sum provided for site investigation precludes any possiblity of a comprehensive survey in which directional trends might be recognised.

In such situations there is no choice for the engineer but to assume spatial uniformity, an assumption that can lead to errors which are not necessarily conservative.

ii) Soil mass has directional trends

There are often occasions when natural soil deposits have parameters that exhibit directional trends, which can range in direction from vertical to horizontal.

With important structures supported on soils with directional trends it may be uneconomical to ignore these trends by thinking only in terms of a mean value and a variance.

Fortunately it is generally a simple matter to fit the measured values to a line of best fit or regression line and to then allow for the random variations of the parameter values about the expected value obtained from this line. For most soil parameters the best regression line is a straight line.

The standard deviation of the material property can then be considered as the square root of the variance about the regression line rather than about the mean value. This value is known as the standard error of the estimate and its use is illustrated below.

Example 8.4

Unconfined compression tests carried out on undisturbed samples taken from a deep clay deposit gave the following results:-

Depth (m)	1	2	3	4	5	6	7	8	9
c_u (kN/m^2)	27	19 21 23	34	37 40 34	42 38	55	48 52	52 58 55	62 58

i) Determine a suitable regression line that will give the relationship between the depth, z, (in metres) and c_u, the strength of the soil at that depth.

ii) A strip foundation, 2m wide, is to be founded at a depth of 1.5m below the surface of the clay. Determine values for the mean and standard deviation of c_u that could be used to evaluate the probability of bearing capacity failure.

Solution

A scatter diagram of the measured values of c_u is shown in Fig.8.4 together with a possible regression line.

<u>Determination of linear regression line formula</u>

A general equation for the regression line is:-

$$c_u = A + Bz$$

where A and B are numerical constants that can be determined in a variety of ways. Probably the most popular methods for hand calculations are those based on the method of least squares. However as most programmable calculators now have suitable subroutines, the values of A and B can be found with a minimum of effort.

Hence A = 15.80 and B = 5.006

The equation of the regression line is therefore

$$c_u = 5.006z + 15.8\text{kN/m}^2 \text{ (where z is in metres)}$$

The variance of c_u about this line is given by:-

$$s^2 = \frac{\sum \left(c_u - c_u(e)\right)^2}{N - 2}$$

Where c_u = measured values of c_u

$c_u(e)$ = expected value of c_u as obtained from the regresssion line formula.

It is seen that the denominator of the formula uses N - 2 rather than N, the total number of measurements. This is general practice in statistics as it produces an unbiased estimate of s, Benjamin and Cornell, (1970).

s, the square root of the above expression, is known as the standard error of the estimate and is analagous to the standard deviation of a single random variable. It does not have a constant value and varies over the depth of the soil layer, (the value calculated by the above formula is the average value of s).

The variance of s, σ_s^2, $= \frac{s^2}{N}[1 + \frac{(z - m_z)}{\sigma_z^2}]^2$

It can be seen that σ_s^2 varies with the value of z.

Design values for the mean and standard deviation of c_u

Obviously the values chosen for the mean and standard deviation of c_u must be representative of the soil contained within the zone of influence of the foundation. For a homogeneous deposit the values of the soil variables at a depth B below the foundation (where B = foundation width or diameter) are generally taken as representative of the deposit.

In the example the foundation is 2m wide and is to be founded at a depth of 1.5m. The representative depth, z, is therefore equal to 1.5 + 2 = 3.5m.

The calculations are probably best set out in tabular form:-

z	c_u	$c_u(e)$	$[c_u - c_u(e)]^2$
1	27	20.81	38.32
2	19	25.81	46.38
2	21	25.81	23.14
2	23	25.81	7.90
3	34	30.82	10.11
4	37	35.83	1.37
4	40	35.83	17.39
4	34	35.83	3.35
5	42	40.83	1.37
5	38	40.83	8.01

6	55	45.84	83.91
7	48	50.84	8.07
7	52	50.84	1.35
8	52	55.85	14.82
8	58	55.85	4.62
8	55	55.85	0.72
9	62	60.85	1.32
9	58	60.85	8.12
			Σ280.27

$$s^2 = \frac{280.27}{18 - 2} = 17.52$$

$$\sigma_s = \sqrt{\frac{17.52}{18}\left[1 + \frac{(3.5 - 5.22)}{6.89}\right]^2} = 0.74$$

From the regression line equation:-

For $z = 3.5$m; $m_{cu} = 33.32$kN/m^2

and $\sigma_{cu} = \sqrt{17.52} + 0.74 = 4.93$kN/m^2

These are therefore the values for the mean and standard deviation of c_u for a Level II reliability analysis of the bearing capacity of the foundation, taking c_u as being normally distributed about the regression line.

If the measured values of c_u had been assumed to be of a purely random nature with no directional trend then the designer would have used a mean value of 41.94kN/m^2 for c_u and a standard deviation of 6.89kN/m^2.

If we assume that the mean and standard deviation values of the soil's unit weight, γ, and the applied uniform pressure, p, were:-

Variable	Mean	Standard deviation
γ	20	1 kN/m^3
p	115	15 kN/m^2

then it is possible to compare the effects of the two approaches:-

Allowing for directional trends gives a reliablity index of 3.57, corresponding to $P_f = 1.8\times10^{-4}$, and F, the factor of safety based on mean values, = 2.43.

Assuming purely random values of c_u gives a reliability index equal to 3.86 ($P_f = 5.6 \times 10^{-5}$) and F = 3.06.

Correlation

Two variables are said to be correlated when the value of one affects the value of the other. If, for instance the angle of shearing resistance of a granular soil increases with the density of the soil then we say that the angle of friction and the density are correlated.

Correlated variables are often referred to as joint variables.

Strictly speaking directional trends, as illustrated in Example 8.1, are a form of correlation but are usually dealt with along the lines shown in the example.

Covariance

The variance of a single random variable, X, is the second moment of area of the pdf diagram (= 1.0) about the mean value:-

$$\sigma^2 = \frac{\sum (x - m_X)^2}{n}$$

The variance of two joint variables, X and Y is COV(X,Y)

where $\mathrm{COV}(X,Y) = \sum \frac{(x - m_X)(y - m_Y)}{n}$

The formula is more convenient when written as:-

$$\mathrm{COV}(X,Y) = \frac{1}{n}\left| \sum xy - \frac{\sum x \sum y}{n} \right|$$

The standard error of the estimate, s , or more correctly $s_{X,Y}$, is of course the square root of COV(X,Y).

The linear correlation coefficient

This important term, given the symbol r, is a means of expressing numerically just how closely the regression line relating the joint variables X and Y fits the observed values and is called the linear correlation coefficient or the degree of correlation.

The degree of correlation, r, is a normalised version of the covariance and is obtained from the formula:-

$$r = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

r can be either positive or negative and vary in value from -1 to +1. If r is negative there is a negative correlation and Y will decrease in value as X increases.

If Y increases as X increases then there is a positive correlation between X and Y.

If there is perfect correlation then for any given value of x the value of y estimated from the regression line formula, y(e), will be equal to the observed value, y.

Alternatively if there is no correlation between X and Y then r = 0.

The following is a rough guide as to the interpretation of the numerical value of r:-

$\|r\| \geq 0.8$	Strong correlation betwwen X and Y which can be assumed to be completely dependent.
$0.8 > \|r\| > 0.2$	Correlation between X and Y
$\|r\| \leq 0.2$	Weak correlation between X and Y which can be assumed independent of each other.

Note

We are only dealing with linear correlation in these notes. An $|r|$ value less than 0.2 indicates little linear correlation between X and Y but it may well be that there is a strong correlation of some non-linear form, a situation that cannot be discussed here.

Treatment of correlated variables

The Level II method of reliability analysis is based on the assumption that the variables in the limit state equation are statistically independent, i.e. they are not correlated.

The procedure for allowing for correlation effects is to use the covariance matrix to transform the space of the corrrelated variables into a space where there is no correlation between them.

The covariance matrix is simply a matrix which lists the correlations between the correlated variables. The leading diagonal is made up of the values of the variances of the correlated variables and the off-diagonal terms are the corresponding covariances.

The space transformation is achieved by an orthogonal transformation in which the axes defining the original variables are lined up parallel with the eigenvectors of the covariance matrix.

The procedure is identical to that used for the determination of principal stresses given the normal and shear stress values that are acting on a set of orthogonal planes. Examples of this procedure have been prepared by the writer, Smith, (1971).

Essentially the method consists of determining the eigenvalues and eigenvectors of the covariance matrix which are, respectively, the variances and the mean values of a set of equivalent, but transformed and uncorrelated, variables.

The method seems to work satisfactorily with structural problems, Hasofer & Lind, (1974), but the writer has had little practical success when attempting to apply it to geotechnical problems.

A solution involving the use of conditional distributions has been proposed by Hohenbichler and Rackwitz, (1981) but the procedure is cumbersome and it must be rmembered that information regarding conditional distributions of soil variables will always be virtually non-existent.

Kiureghian and Taylor, (1983) report that a heuristic solution has been proposed by Kitagawa and Kiereghian, (1980), but acknowledge that the accuracy of the procedure has not yet been adequately examined.

However the dominant correlation in any soil strength problem is the relationship between density and the angle of shearing resistance and it may be that the linear regression analysis described in the earlier part of this chapter can be used to allow for this.

The following example is of interest and is presented to the reader as a possible solution to the problem of soil correlations.

<u>Example 5.1</u>

A series of shear tests carried out on a set of random samples taken from a sandy soil deposit gave the values of unit weight and corresponding peak angle of drained shearing resistance listed below.

Unit wt.(kN/m^3)	18.0	18.3	18.6	18.8	19.0	19.1	19.2	19.2	19.3	19.4
ϕ (degrees)	33.0	33.1	33.3	33.5	33.3	33.6	33.8	34.2	34.5	34.8
Unit wt.(kN/m^3)	19.5	19.5	19.6	19.7	19.9	20.0	20.1	20.2	20.2	20.2
ϕ (degrees)	35.0	35.2	35.4	35.5	36.2	36.5	36.8	37.1	37.0	37.2

Using these soil properties determine the reliability index for the bearing capacity problem of example 1.6:-

i) Assuming no correlation between γ and ϕ.

ii) Allowing for the correlation between γ and ϕ.

Solution

The values of the mean and s.d. for the soil variables are:-

Unit weight - mean = 19.39kN/m^3; s.d. = 0.636kN/m^2

Friction angle - mean = 34.95^o; s.d. = 1.46^o

Let X = γ and Y = ϕ

It can quickly be found that:-

$\Sigma x = 699$; $\Sigma y = 387.8$; $\Sigma xy = 13570.4$

$$\text{Hence COV}(X,Y) = \frac{1}{20}\left[13570.4 - \frac{699 \times 387.8}{20}\right]$$

$$= 0.837$$

$$\Rightarrow \quad r = \frac{0.837}{1.46 \times 0.636} = 0.901$$

This value of r shows that there is actually a strong positive correlataion between γ and ϕ.

The regression line formula that relates the unit weight and the angle of shearing resistance is found to be:-

$$\gamma = 4.9618 + 0.4128\phi$$

and that $s^2 = 0.0428$ and $\sigma_s = 0.2068$

If we consider the expected value of γ corresponding to the mean of ϕ (34.95^o) then $\sigma_\gamma = 0.2068 + \sqrt{\frac{.0428}{18}} = 0.256$

i) Assuming no correlation

Basic variables are:-

Variable	Symbol	Mean	s.d
γ	X_1	19.39	0.636
N_γ	X_2	33.66	8.13
P	X_3	500	30

The suggested method gives $\beta = 0.872$

ii) Allowing for correlation.

The basic variables are as above except that the s.d. of the unit weight is reduced from 0.636 to 0.256.

The suggested method gives $\beta = 0.876$

Note

Monte Carlo simulations of the problem gave $\beta = 1.099$ with no allowance for correlation and $\beta = 1.153$ when correlation was allowed for.

Whilst firm conclusions cannot be drawn from a single example it is interesting to note that there are indications that the correlation that undoubtedly exists between soil density and soil strength may not prove to be of great significance in reliability analysis work.

There are also indications that the value of P_f determined when correlation effects are ignored, whilst being conservative, is not far removed from the value of P_f when correlation is allowed for.

Summary

In this chapter an attempt has been made to list some of the uncertainties that are connected with the measured material properties of soils.

The main reason for these uncertainties is of course the inevitable shortage of samples and because of the financial considerations but another reason is lack of knowledge on testing errors.

In this context the term, "test error", is defined as the difference in the measured test value of a parameter that occurs if the test is carried out twice, using different machines and operators. It is not meant to be the error caused by a gross mistake on the part of an operator.

If operators and testing machines were somehow randomly selected then the test errors would also be random. However, for most site investigations, only one soils testing firm is usually involved, possibly just one operator and just one testing machine.

In this case the test errors will tend to be constant rather than random and will be biased, i.e. they will consistently produce either high or low values.

In order to investigate the effect of the various factors that make us a testing error there is obviously a need for large research programmes in which standard soil samples are sent to different laboratories and tested by different operators on different machines.

Some work has been done along these lines, Hammitt, (1966) and Sherwood, (1970) but much remains to be done.

Lack of information can force us to use the assumption of normal distributions for the basic soil parameters, γ, c and ϕ although guidance for the adoption of more realistic distributions has been given in chapter 5.

For related variables, such as $\tan\phi$, N_{γ}, etc. the form of their distributions can be found, (see chapter 7).

Correlations between variables are particularly difficult to obtain when information is limited. The results from a typical site investigation will generally only be sufficient to detect simple lineal horizontal or vertical directional trends.

With such a lack of statistical knowledge regarding geotechnical uncertainties, (only soil properties have been discussed here), it can be appreciated that the application of the Level III method of reliability analysis to soils engineering is not practical.

The Level II method, i.e. the second moment approach, is obviously more suitable for soils.

The proposed adaption of the second moment method, as described in this thesis, appears to create a procedure robust enough to give satisfactory predictions even when the statistical information available is relatively limited.

CHAPTER 9 - POSSIBLE FUTURE DEVELOPMENTS

Recent developments

To make some assessment of fututure developments that could occur in the field of geotechnical reliability analysis it is necessary to look at some of the changes that have taken place in that subject during the recent past.

For several years the established structural design methods used in the United Kingdom have been under considerable pressure for change.

Perhaps the most radical innovation was the switch from imperial to metric units, resulting from Britain's entry into the European Economic Community in 1973.

Further pressure for change came from the interest being shown in the possible use of statistics and probability theory in structural design work, brought on by the rapid advent of programmable calculators and microcomputers.

Recently there has been increasing realisation amongst design engineers that allowing for uncertainties in the structural parameters, rather than assuming constant values and taking large factors of safety, can lead to more economical structures.

In 1972 CP110, Part I, "The structural use of concrete", became the first British Code of Practice to accept that there were serviceablity states other than the ultimate and to advocate the use of limit state design.

The Code was largely a translation of its previous allowable stress editions being along the lines of a Level I design method in that characteristic strengths and loads, to be used with appropriate partial factors of safety, were suggested. Probability theory was not mentioned directly.

The possible adoption of limit state design in Britain was more or less accepted by structural engineers but amongst soils engineers there were grave doubts. Many engineers had mistakenly decided that limit state design is only possible with the use of statistics and probability theory.

A British Geotechnical Society meeting held in January, 1981 highlighted the controversy and a summing up of the situation at that time has been prepared by Boden, (1981).

Eurocode 7

Further pressure for change has come from the European Economic Commission's decision, in 1972, to prepare a set Eurocodes, i.e. codes of practice for building and civil engineering works.

The Commission's plan is to produce 7 Eurocodes:-

1 - Concerning unified rules.

2 - Concerning reinforced concrete structures

3 - Concerning steel structures

4 - Concerning composite (steel and concrete) structures

5 - Concerning wooden structures

6 - Concerning brick structures

7 - Concerning foundations

Supplements, concerning loadings and earthquake effects may also be produced.

Code 1 specifies the definitions to be used in the other codes and lays down the common basis of rules to be specified for each code in order to maintain a uniformity in the codes. Eurocode 1 therefore is not a design code, merely a guide for the preparation of the other Eurocodes.

At the present time Eurocode 1 is now in its third draft whilst codes 2 and 3 are in their first draft. Base documents, from which drafts of codes 4, 6 and 7 will be prepared for discussion, are now almost complete.

Only aspects of Eurocode 7 will be considered here.

In 1980 the late Professor K. Nash, then Secretary General of the International Society for Soil Mechanics and Foundation Engineering, suggested that the Society could prepare a base document for Eurocode 7.

An agreement was reached between the E.E.C. and the I.S.S.M.F.E. in which the latter undertook to carry out a survey of the codes of practice for foundations within the member countries and to prepare a base document that could then be used for the preparation of a draft edition of Eurocode 7.

This document has now been prepared by the committee set up by the I.S.S.M.F.E. which has a representative from each of the European countries involved.

B. Simpson is the U.K. representative on the committee and has prepared a summary of the contents of the base document, (1983). The base document consists of ten chapters most of which are substantially complete, the main exception being the one on loading which has proved particularly difficult to prepare.

The draft of the base document was discussed at a meeting of the British Geotechnical Society in London on May 12th, 1983. This meeting has been reported by Driscoll, (1984).

The writer was at the meeting and, in his opinion, the overall feeling amongst those present was that it have been better if Britain had been allowed to stick to its present foundation code rather than be forced to change to a European code which, because of its wider regional extent, might loose sight of some of the local problems that apply in Britain.

The writer obtained the impression that many of the members present were apprehensive of the document's apparent backing for the use of limit state design because of their own lack of understanding of the principles involved.

During the evening the meeting agreed, by vote, that soil is not a material that can be modelled statistically and that the evaluation of values for characteristic strengths/loads and partial factors of safety by experience and by the application of statistics and probability theory is not applicable to geotechnical engineering.

Britain is only one of nine voices at the I.S.S.M.F.E. Committee and, perhaps not suprisingly, the resolution of the B.G.S. regarding characteristic values was not accepted at the next meeting of the Committee, just two weeks later on May 25th at the European Conference on Soil Mechanics and Foundation Engineering, in Helsinki.

The present position is therefore that the base document for Eurocode 7 is approaching finalisation. The indications are that it will advocate the use of limit state design using Level I approach involving the use of characteristic values and partial factors of safety similar to those used in the present Danish Code of Practice for foundation engineering.

Characteristic value x_k

The characteristic value of a random resistance variable, X, is defined as the value of X below which only p% of possible X values may fall.

The characteristic value of a random loading variable, X, is defined as the value of X above which only p% of possible X values may fall.

p is an arbitrarily chosen value of probability, usually 5%

The general expression for the characteristic value is:-

$$x_k = m_X \pm k\sigma_X$$

k is a multiplier and its value will depend upon the probability distribution of the variable. For a normal distribution k = 1.645.

Future developments

With this continuing impetus for change there is little doubt that there is an increasing need for methods of reliability analysis that can be used by soils engineers to determine the safety of their structures.

It is not hard to make a prediction as to how future work in geotechnical reliability analyses will be concentrated.

For a given structure, subjected to a particular type of loading, and using the methods outlined in this thesis, a nominal probability of failure value, P_f, can be obtained.

By a consideration of the social and economic demands that will be made by the society in which a structure will be constructed it is possible to determine acceptable probability of failure values corresponding to particular limit states of that structure.

The acceptable probability of failure is usually referred to as the "target probability" and given the symbol P_{ft}.

Obviously, for an economical design the nominal probability of failure of the structure, P_f, need not be less than P_{ft}.

The determination of risk levels for the design of different types of geotechnical structures in different locations will be a necessary part of future reliability work.

Some work has already been carried out along these lines, see chapter 4. Possible procedures are described in Report No.63, C.I.R.I.A., (1976).

Once P_{ft} values have been obtained the calculation of partial factors for use with Eurocode 7 can be obtained by a Level II reliability analysis.

Partial factors of safety can then be evaluated for the relevant parameters, either in terms of mean values:-

$$\gamma_i = \frac{x_i}{m_{x_i}}$$

or in terms of characteristic values:-

$$\gamma_{k,i} = \frac{x_i}{x_{k,i}}$$

where x_i is the operating value of the parameter x at failure. These operating values are often called, "design values".

Obviously characteristic values are multiplied or divided by the partial factors in order to decrease the value of the resistance and to increase the value of the applied loading.

Conversely once a set of partial factors is available for a particular structure and knowing the characteristic values it becomes a simple matter to determine values for x_i which, when used in the design calculations, will design a structure with a P_f value equal to P_{ft}.

The possible final form of Eurocode 7

It is expected that design by Eurocode 7 will be much along the lines just described but not for some time. In 1981 Ovesen, the chairman of the I.S.S.M.F.E. Committee, estimated that the code will not be available for the practising engineer before 1986 or even perhaps 1990.

However this writer wonders just how effective the system will eventually prove to be. It might be simpler to use the Level II method directly as a design procedure and not to become involved with partial factors at all.

If the required sets of P_{ft} values for different structures are eventually obtained then obviously sets of acceptable β values will also have been prepared.

The safety of a structure could therefore be checked by simply comparing the β value obtained from a reliability analysis with the value of β that is deemed suitable for the structure.

The reliablity analysis method proposed in this thesis is more than capable of performing this task and a possible further development would be to alter the algorithm so that the structure could be designed directly to have a β value equal to the acceptable value.

With such a procedure the most economical form of the structure would be created.

It will be interesting to see just what form the design procedure adopted by Eurocode 7 will finally take.

Other topics for future work

There are many other aspects of geotechnical reliablity analysis that require further research. Examples are such items as soil and load distributions , correlations, progressive failure in earth slopes and settlement predictions.

The method of reliability analysis proposed in this thesis appears to have the potential to tackle quite complicated soil problems and the writer hopes to use it in the investigation of at least some of the subjects mentioned above.

For settlement problems one cannot think in terms of loads and resistances and the limit state equation will have to be expressed in terms of capacity and demand where capacity will be the allowable settlement and demand will be the predicted settlement value.

The estimation of how the applied loading will affect the settlement is also a problem that requiries investigation.

For a stability problem maximum load values must be used in the analysis but it is recognised that, for a structure subjected to more than one time dependent load, the chance of any two of these loads acting simultaneously at their maximum values is extremely unlikely.

Consequently some economies are often obtained by not designing the structure to carry the sum of all the maximum loads, Turkstra, (1970) and Ferry Borges and Castanheta, (1974), It might be possible to devise a similar form of estimation for the value of load combination that will predominate during the process of long term settlement.

Such a procedure would most likely be based on a probability approach and could prove of great value in settlement predictions.

Bearing in mind that settlement also involves the position of the water table and the permeability of the soil it is seen that the problem of soil settlement is huge and will obviously demand the attention of research workers for many years yet.

APPENDIX I

NUMERICAL VALUES OF N_γ AND ITS FIRST DERIVATIVE

ϕ (Degrees)	N_γ	$\partial N_\gamma/\partial\phi$
0.00	0.00	0.00
1.00	0.00	0.29
2.00	0.01	0.62
3.00	0.02	1.00
4.00	0.05	1.43
5.00	0.07	1.92
6.00	0.11	2.49
7.00	0.16	3.14
8.00	0.22	3.88
9.00	0.30	4.74
10.00	0.39	5.72
11.00	0.50	6.85
12.00	0.63	8.15
13.00	0.78	9.64
14.00	0.97	11.36
15.00	1.18	13.34
16.00	1.43	15.63
17.00	1.73	18.28
18.00	2.08	21.35
19.00	2.48	24.91
20.00	2.95	29.04
21.00	3.50	33.85
22.00	4.13	39.45
23.00	4.88	45.99
24.00	5.75	53.65
25.00	6.76	62.63
26.00	7.94	73.18
27.00	9.32	85.61
28.00	10.94	100.30
29.00	12.84	117.70
30.00	15.07	138.36
31.00	17.69	162.97
32.00	20.79	192.38
33.00	24.44	227.65
34.00	28.77	270.07
35.00	33.92	321.31
36.00	40.05	383.43
37.00	47.38	459.03
38.00	56.17	551.46
39.00	66.76	664.98
40.00	79.54	805.05
41.00	95.05	978.78
42.00	113.96	1195.41
43.00	137.10	1467.08
44.00	165.58	1809.82
45.00	200.81	2245.00
46.00	244.65	2801.29
47.00	299.52	3517.53
48.00	368.67	4446.79
49.00	456.40	5662.28
50.00	568.57	7266.03

APPENDIX II

NUMERICAL VALUES OF N_q AND ITS FIRST DERIVATIVE

ϕ (Degrees)	N_q	$\partial N_q/\partial\phi$
0.00	1.00	5.14
1.00	1.09	5.63
2.00	1.20	6.16
3.00	1.31	6.75
4.00	1.43	7.39
5.00	1.57	8.11
6.00	1.72	8.90
7.00	1.88	9.78
8.00	2.06	10.75
9.00	2.25	11.83
10.00	2.47	13.02
11.00	2.71	14.36
12.00	2.97	15.84
13.00	3.26	17.50
14.00	3.59	19.36
15.00	3.94	21.43
16.00	4.34	23.76
17.00	4.77	26.37
18.00	5.26	29.32
19.00	5.80	32.64
20.00	6.40	36.39
21.00	7.07	40.63
22.00	7.82	45.45
23.00	8.66	50.93
24.00	9.60	57.17
25.00	10.66	64.31
26.00	11.85	72.48
27.00	13.20	81.86
28.00	14.72	92.66
29.00	16.44	105.13
30.00	18.40	119.57
31.00	20.63	136.35
32.00	23.18	155.90
33.00	26.09	178.76
34.00	29.44	205.59
35.00	33.30	237.18
36.00	37.75	274.54
37.00	42.92	318.89
38.00	48.93	371.76
39.00	55.96	435.08
40.00	64.20	511.27
41.00	73.90	603.41
42.00	85.37	715.42
43.00	99.01	852.33
44.00	115.13	1020.66
45.00	134.87	1228.92
46.00	158.50	1488.25
47.00	187.21	1813.45
48.00	222.30	2224.23
49.00	265.50	2747.24
50.00	319.06	3418.69

APPENDIX III

NUMERICAL VALUES OF N_c AND ITS FIRST DERIVATIVE

ϕ (Degrees)	N_c	$\partial N_c/\partial\phi$
0.00	5.14	12.80
1.00	5.38	14.03
2.00	5.63	14.90
3.00	5.90	15.84
4.00	6.19	16.86
5.00	6.49	17.96
6.00	6.81	19.16
7.00	7.16	20.46
8.00	7.53	21.87
9.00	7.92	23.40
10.00	8.34	25.07
11.00	8.80	26.89
12.00	9.28	28.88
13.00	9.81	31.06
14.00	10.37	33.45
15.00	10.98	36.07
16.00	11.63	38.96
17.00	12.34	42.14
18.00	13.10	45.64
19.00	13.93	49.52
20.00	14.83	53.82
21.00	15.81	58.59
22.00	16.88	63.89
23.00	18.05	69.80
24.00	19.32	76.41
25.00	20.72	83.81
26.00	22.25	92.12
27.00	23.94	101.47
28.00	25.80	112.02
29.00	27.86	123.96
30.00	30.14	137.50
31.00	32.67	152.92
32.00	35.49	170.52
33.00	38.64	190.68
34.00	42.16	213.85
35.00	46.12	240.56
36.00	50.59	271.49
37.00	55.63	307.43
38.00	61.35	349.37
39.00	67.87	398.51
40.00	75.31	456.36
41.00	83.86	524.78
42.00	93.71	606.11
43.00	105.11	703.28
44.00	118.37	820.04
45.00	133.87	961.17
46.00	152.10	1132.81
47.00	173.64	1342.94
48.00	199.26	1602.00
49.00	229.92	1923.77
50.00	266.88	2326.62

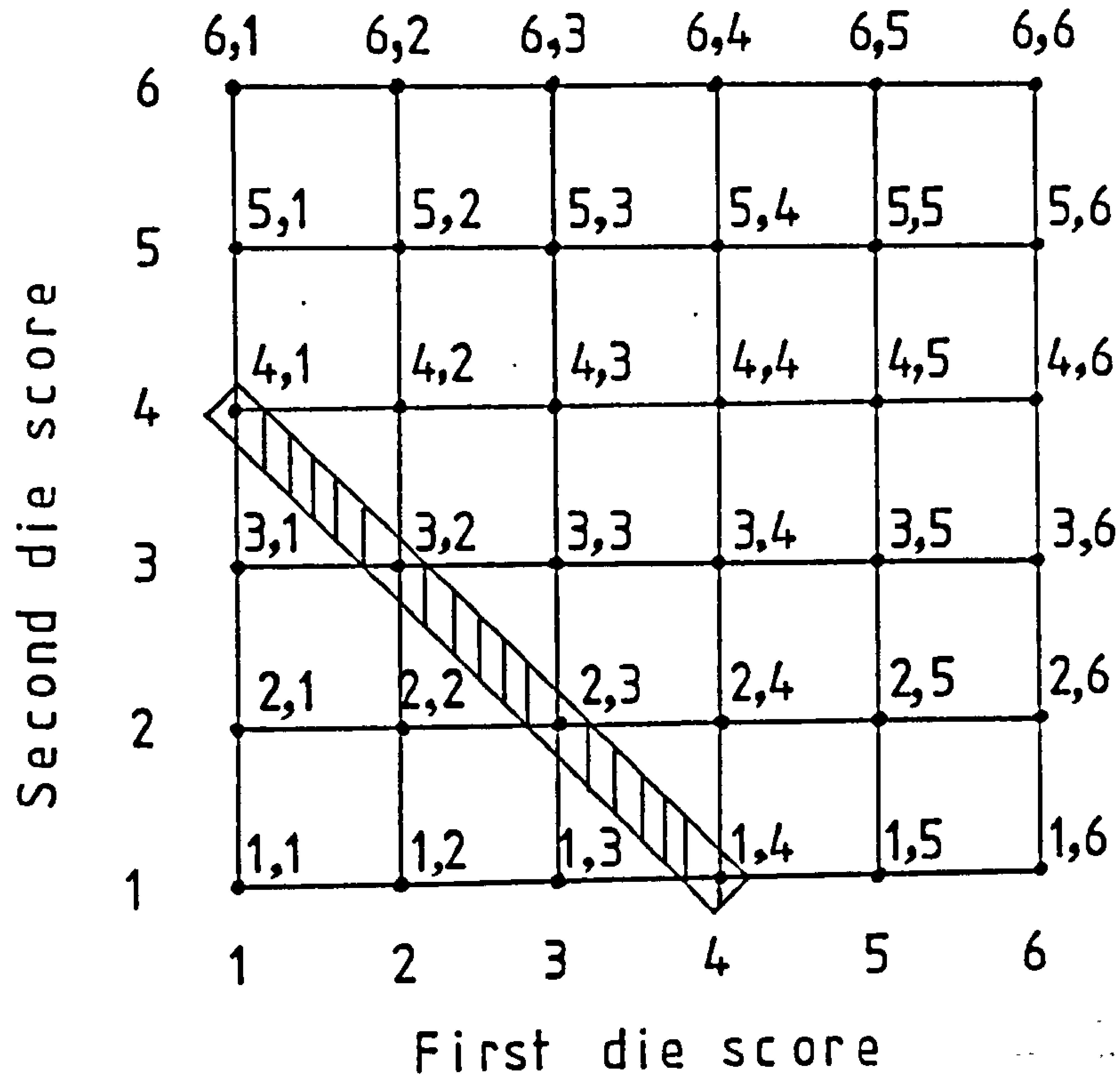


Fig.1.1 Sample space for score of two dice

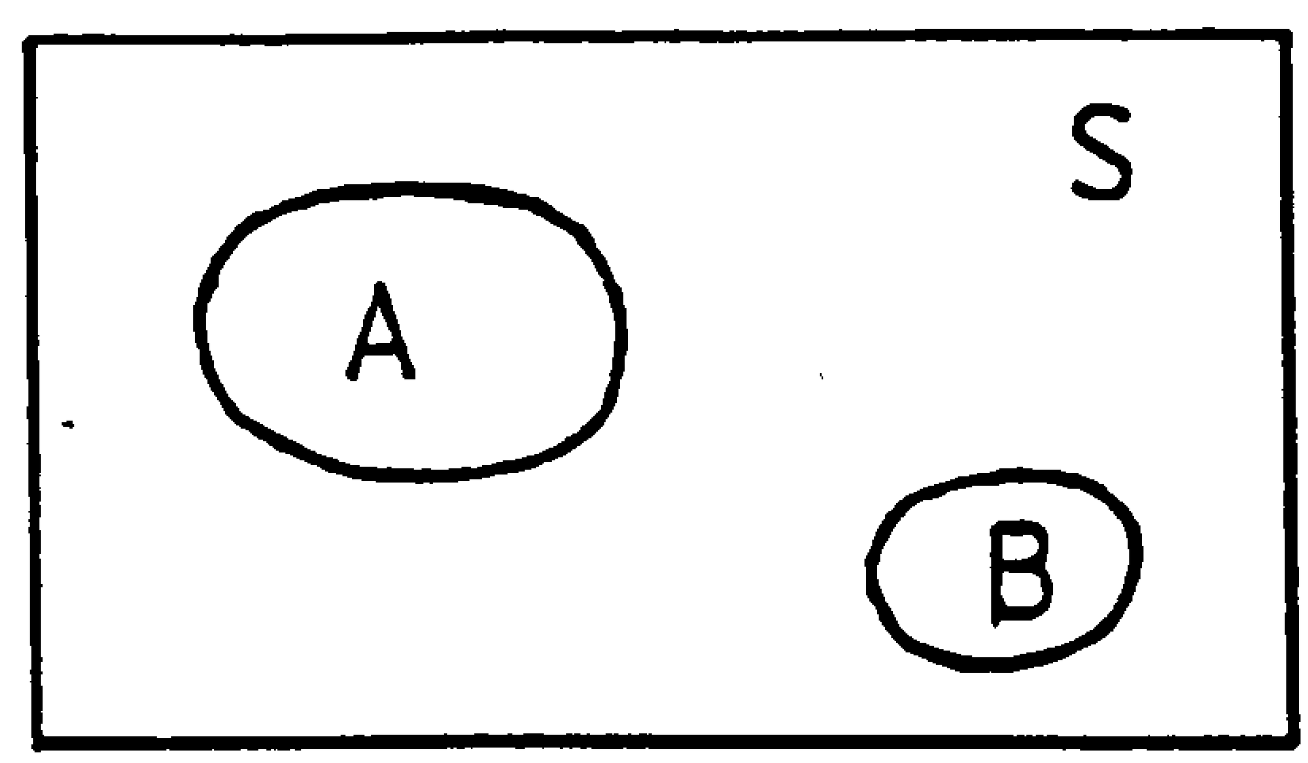


Fig.1.2 Universe S and subsets A and B

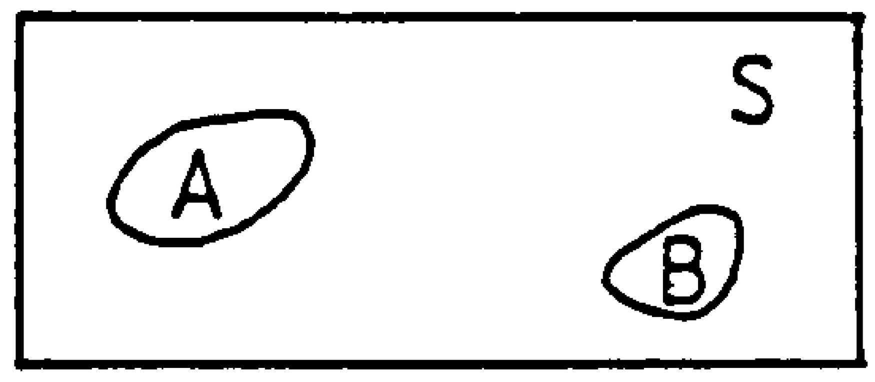
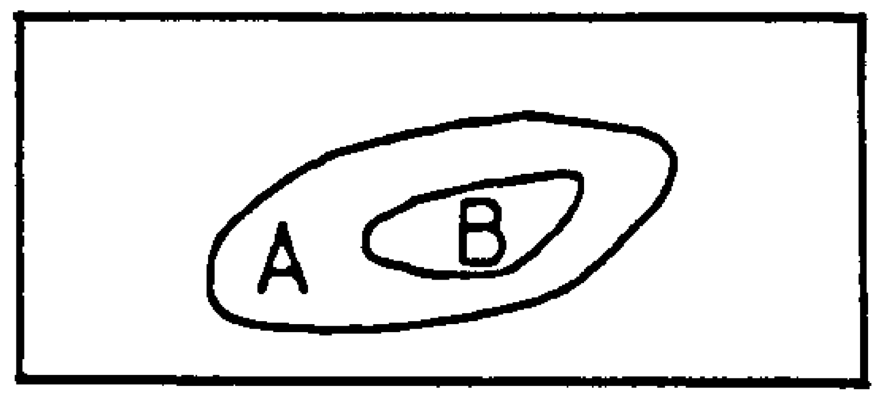
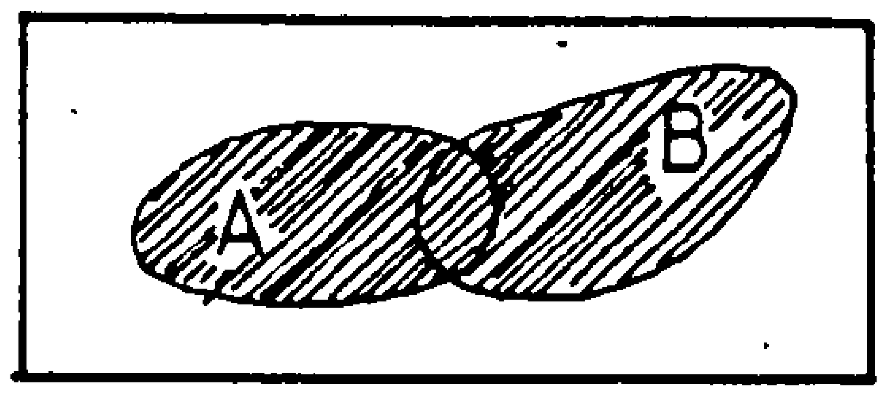
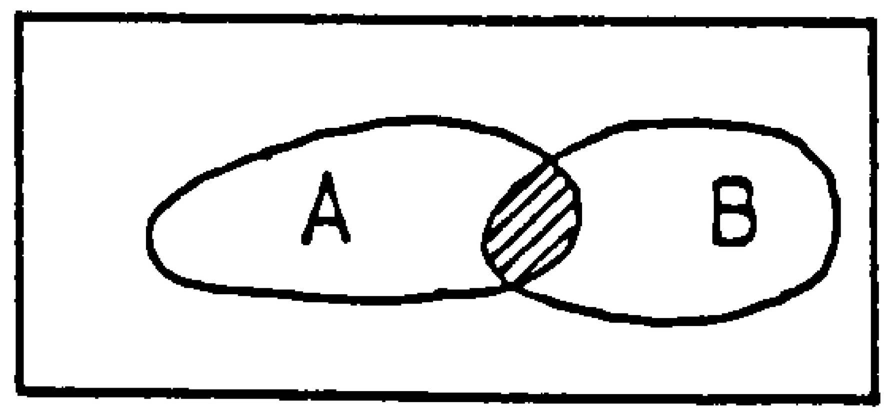
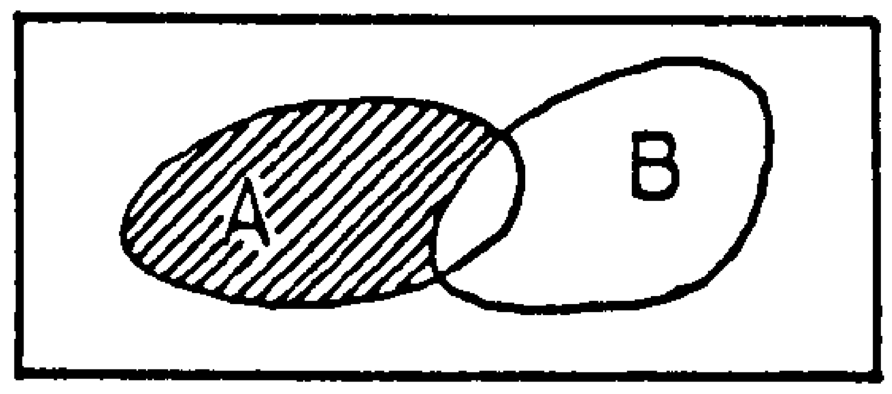
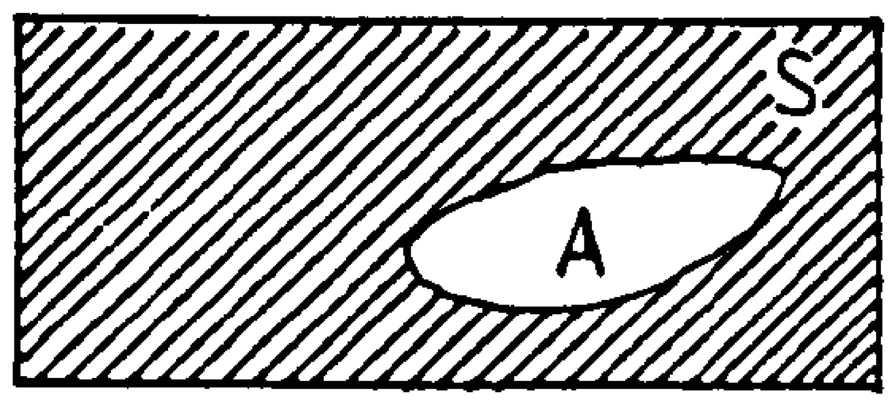
Description	Mathematical expression	Venn diagram
<u>Mutually exclusive events A and B</u> (no common elements)	$A \cap B = 0$	
<u>B is a subset of A</u> (all elements of B are included in A)	$B \subset A$	
<u>Union of A and B</u> (all elements that are in either A or B)	$A \cup B$	
<u>Intersection of A and B</u> (all elements in both A and B)	$A \cap B$	
<u>Difference between A and B</u> (elements in A but not in B)	$A - B$	
<u>Complementary set $\bar{A}$</u> (elements not in A)	$\bar{A} = S - A$	

Fig.1.3 Set operations

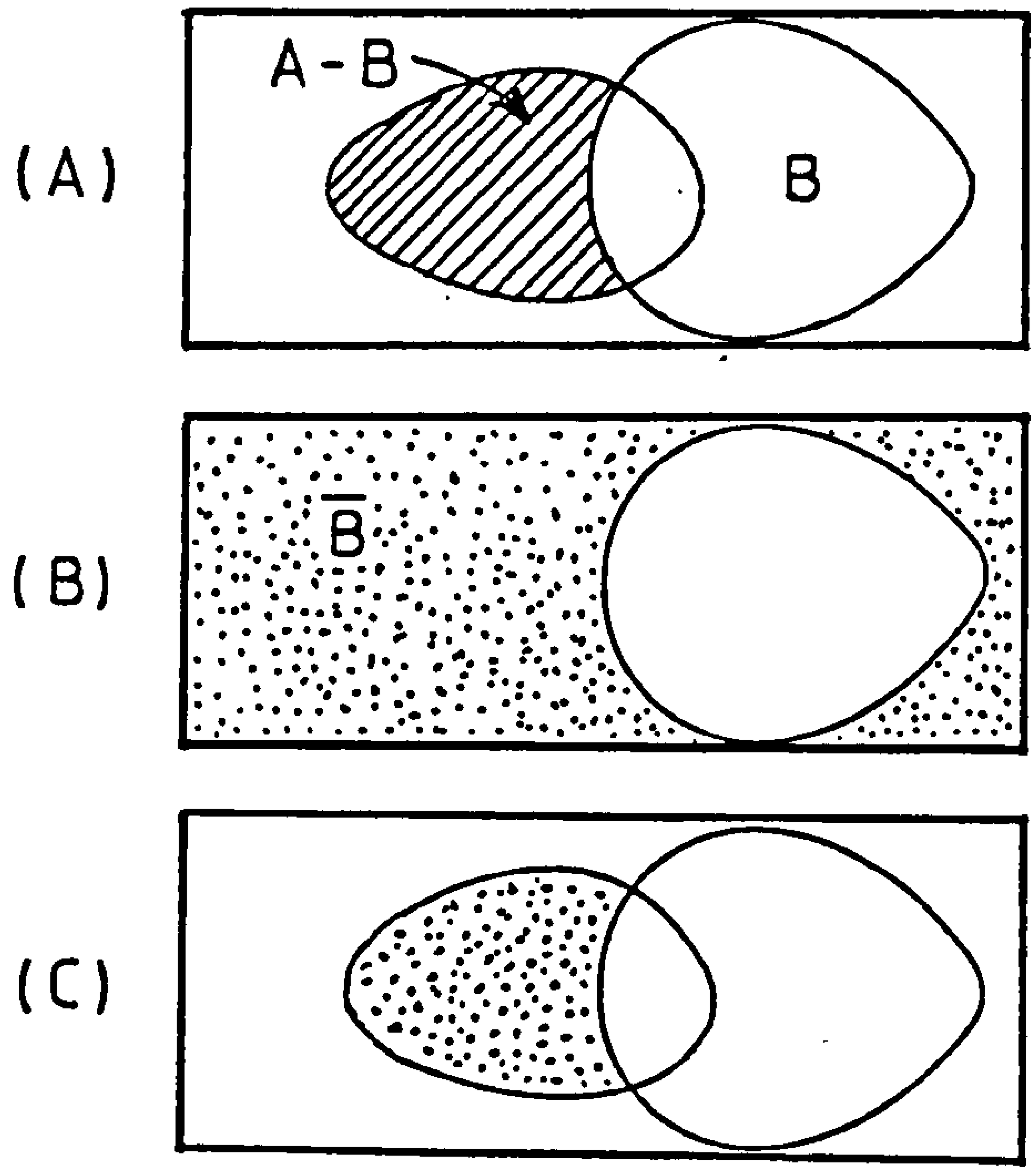


Fig.1.4 Example 1.1

B1+B2+B3+ ······Bn = S

(A∩B1)+(A∩B2)+(A∩B3)+······(A∩Bn) = A

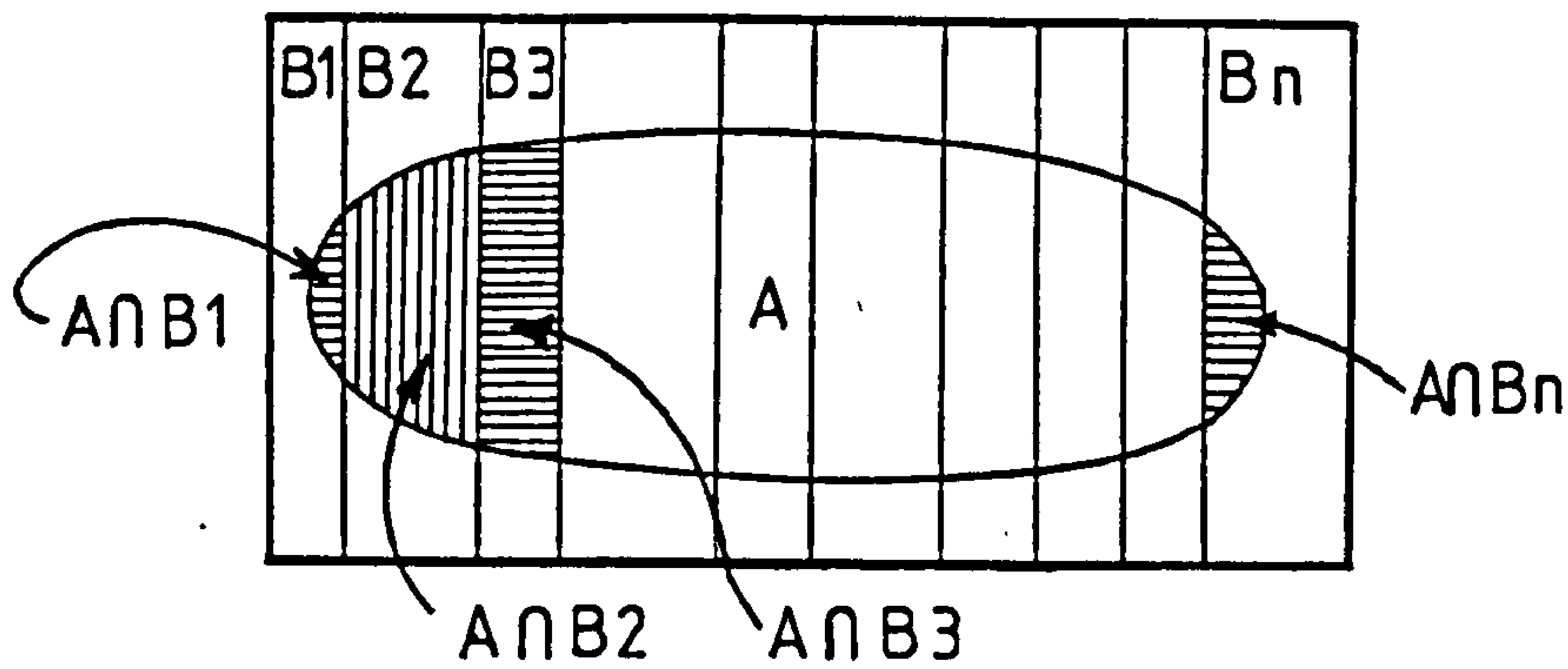


Fig.1.5 Theorem of total probability

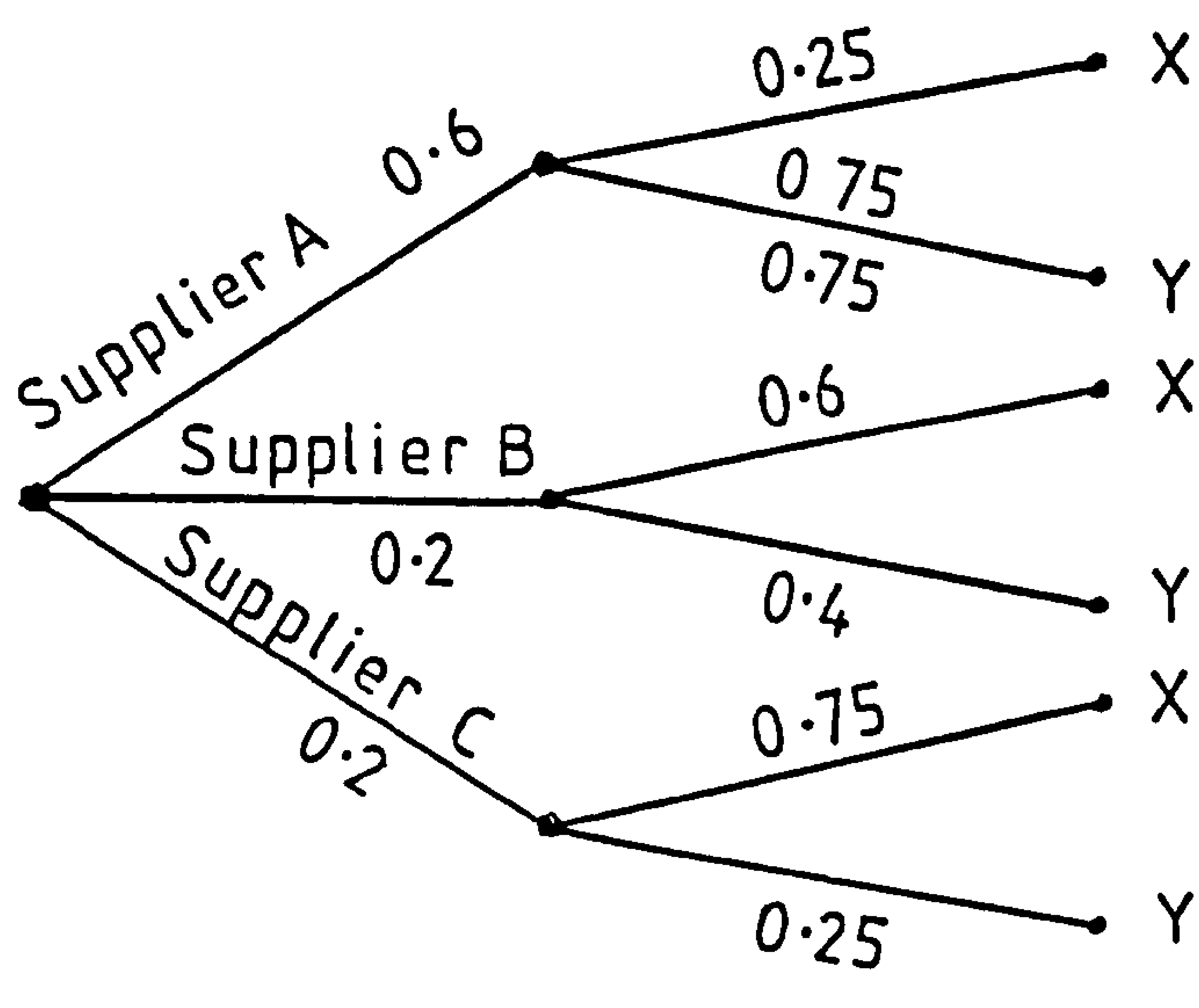


Fig.1.6 Example 1.9

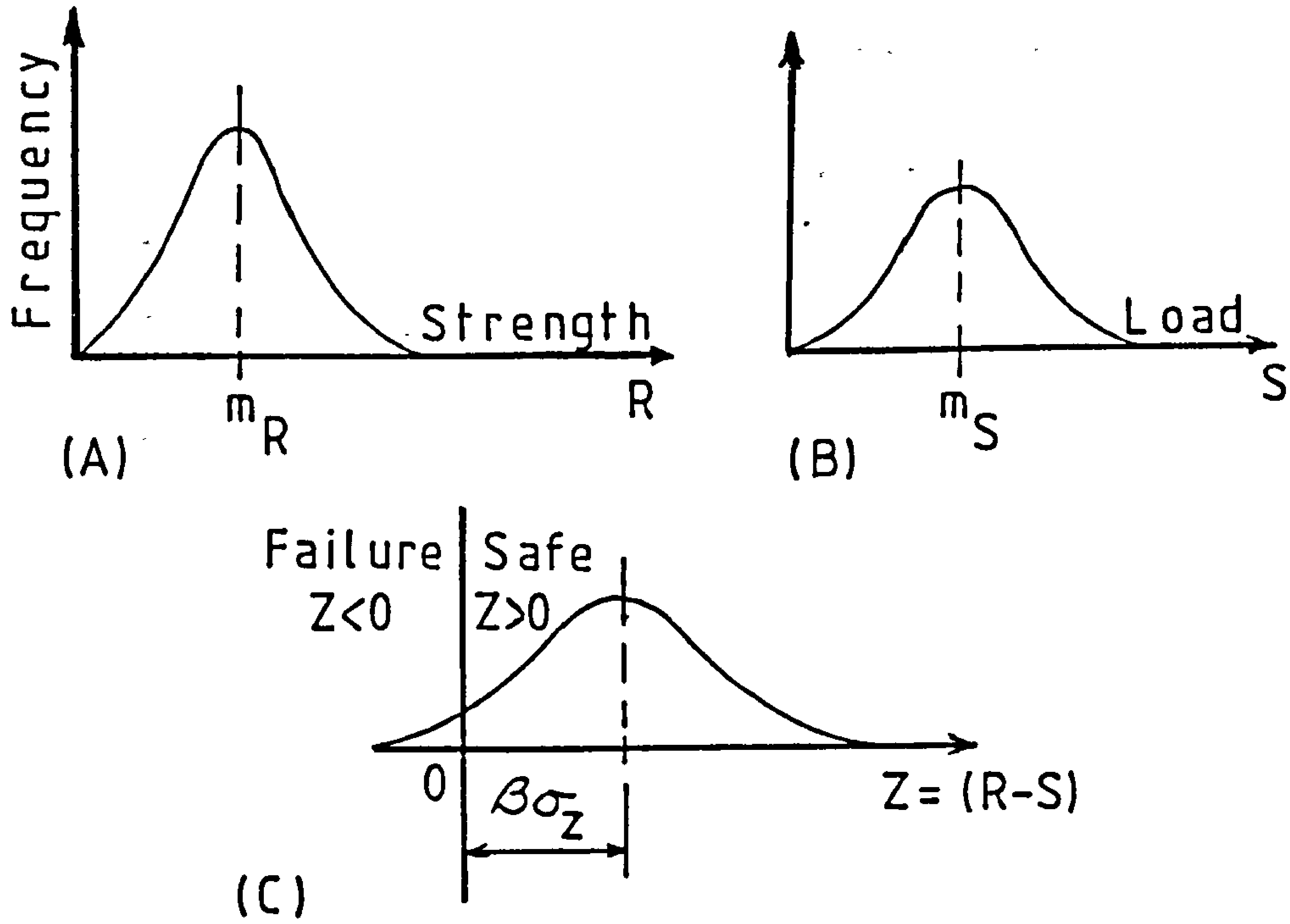


Fig.2.1 The reliability index β

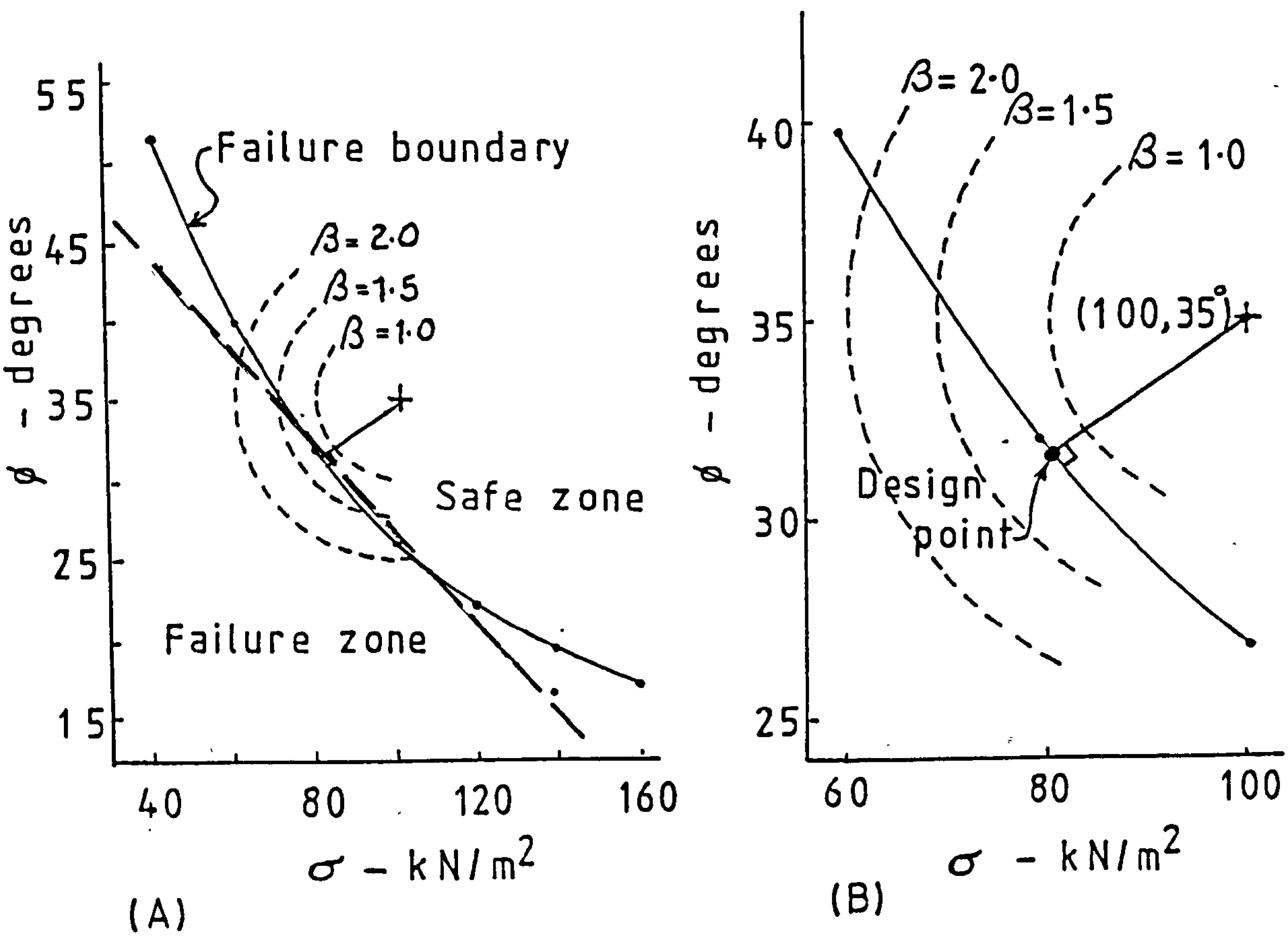


Fig. 2.2 Example 2.1

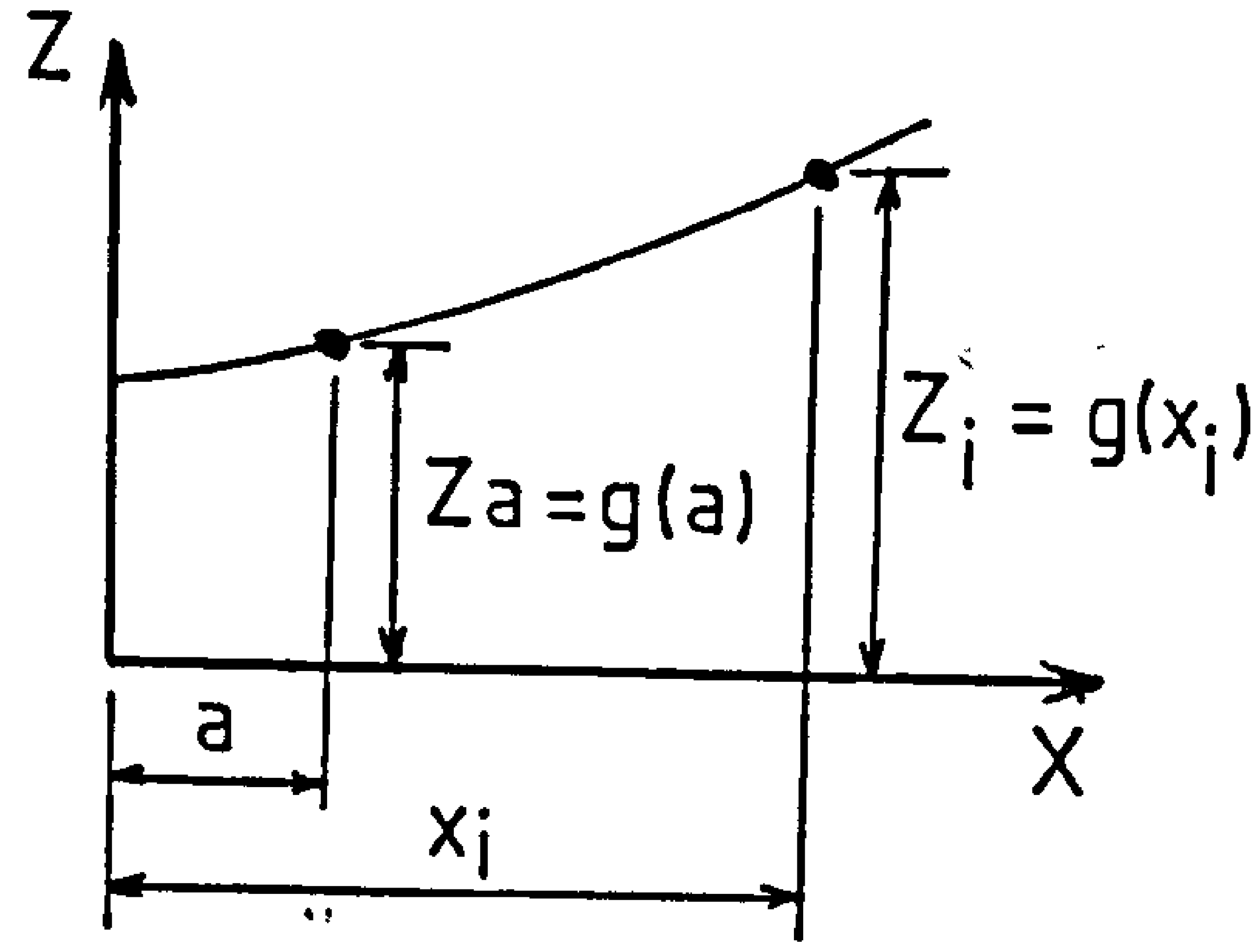


Fig. 2.3

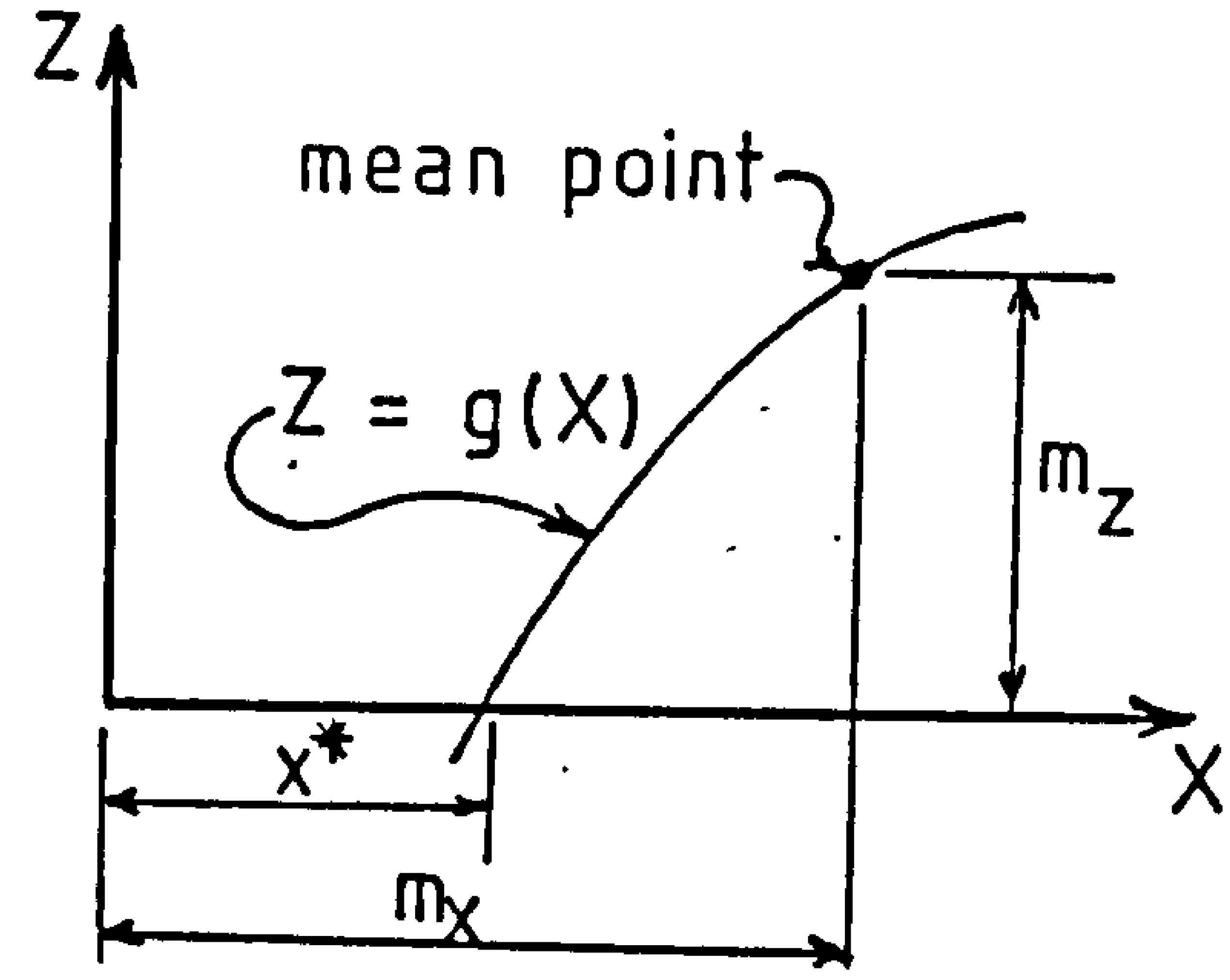


Fig. 2.4

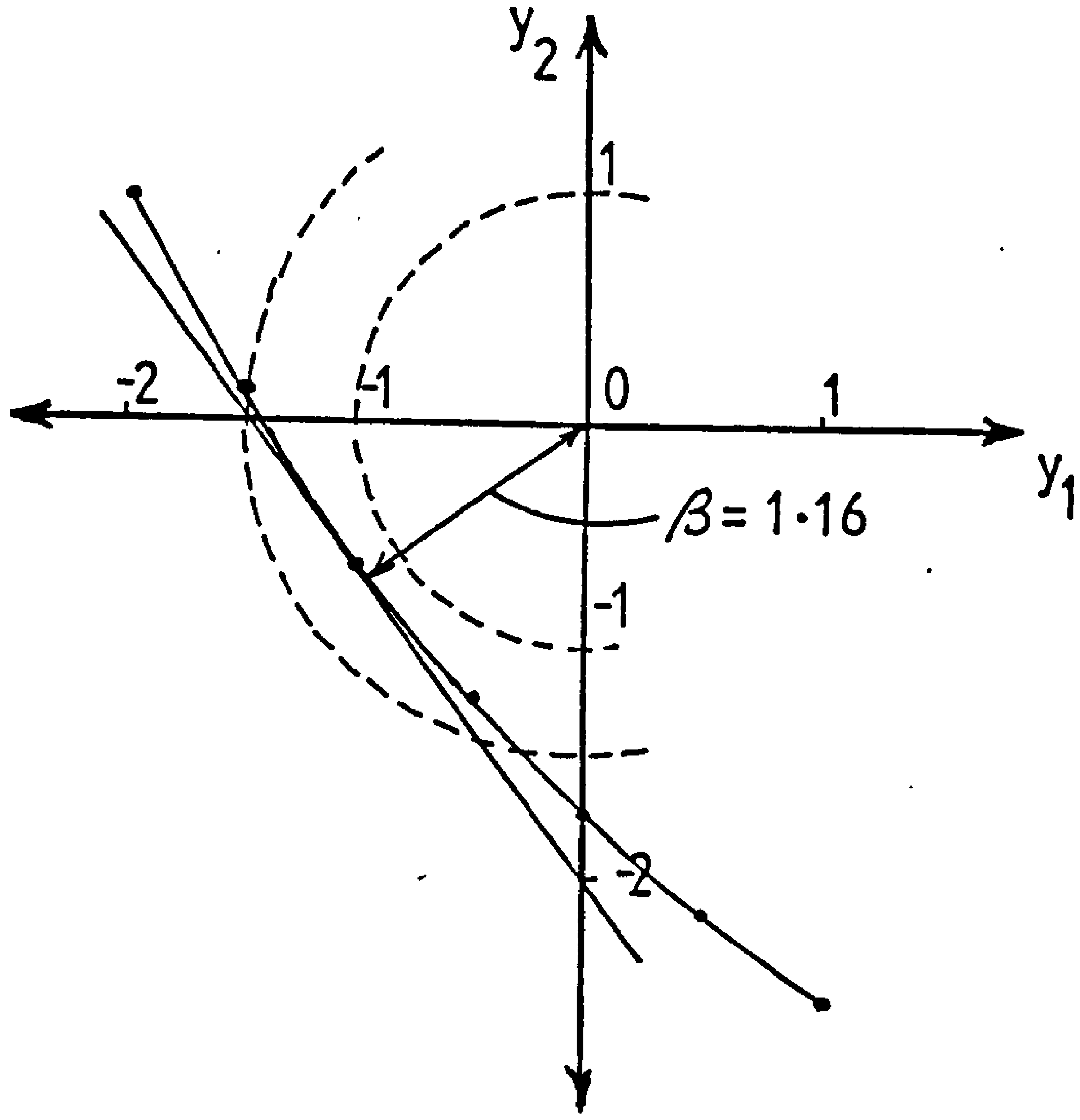


Fig. 2.5 Example 2.6

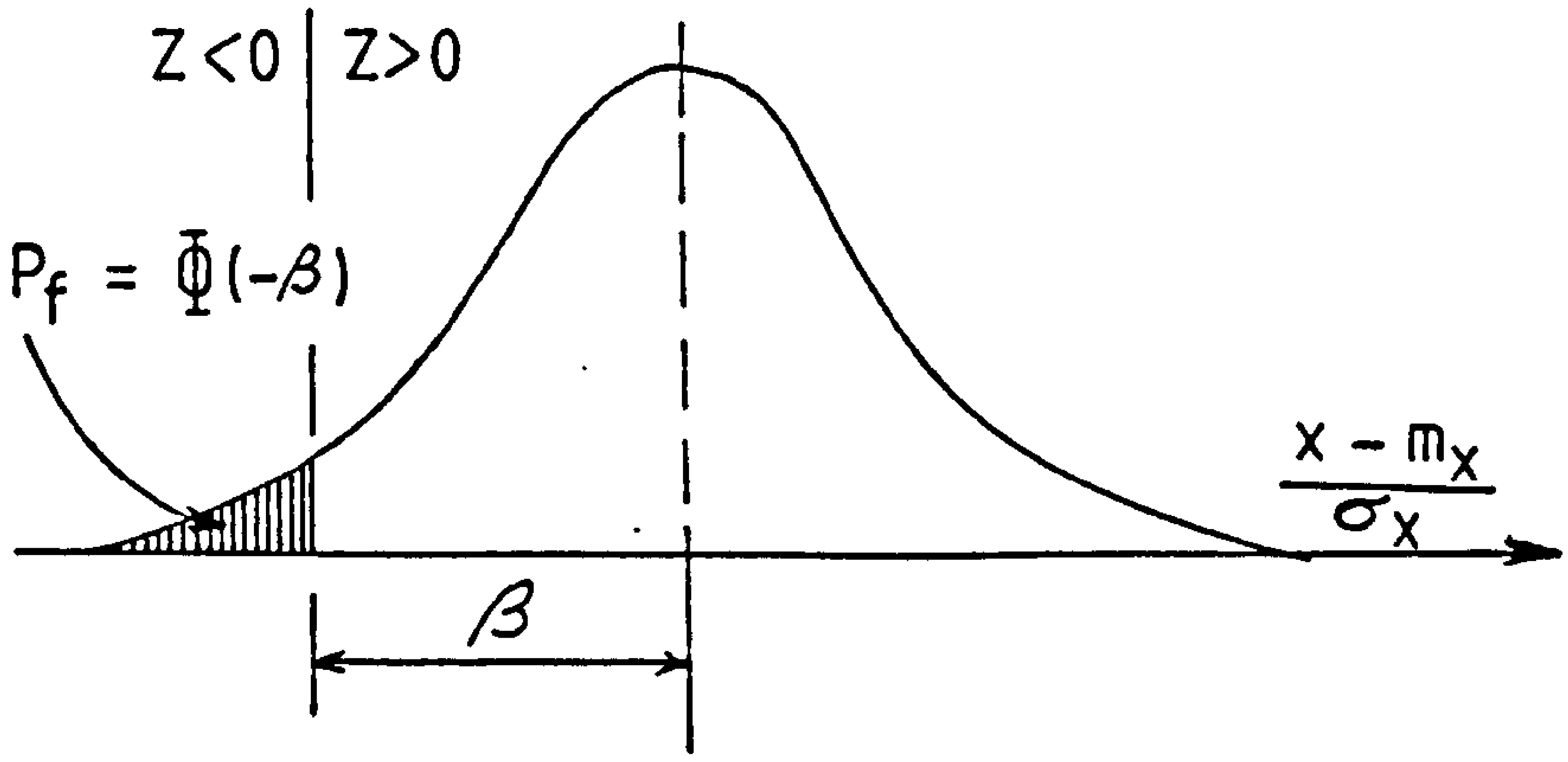


Fig. 2.6 Probability of failure

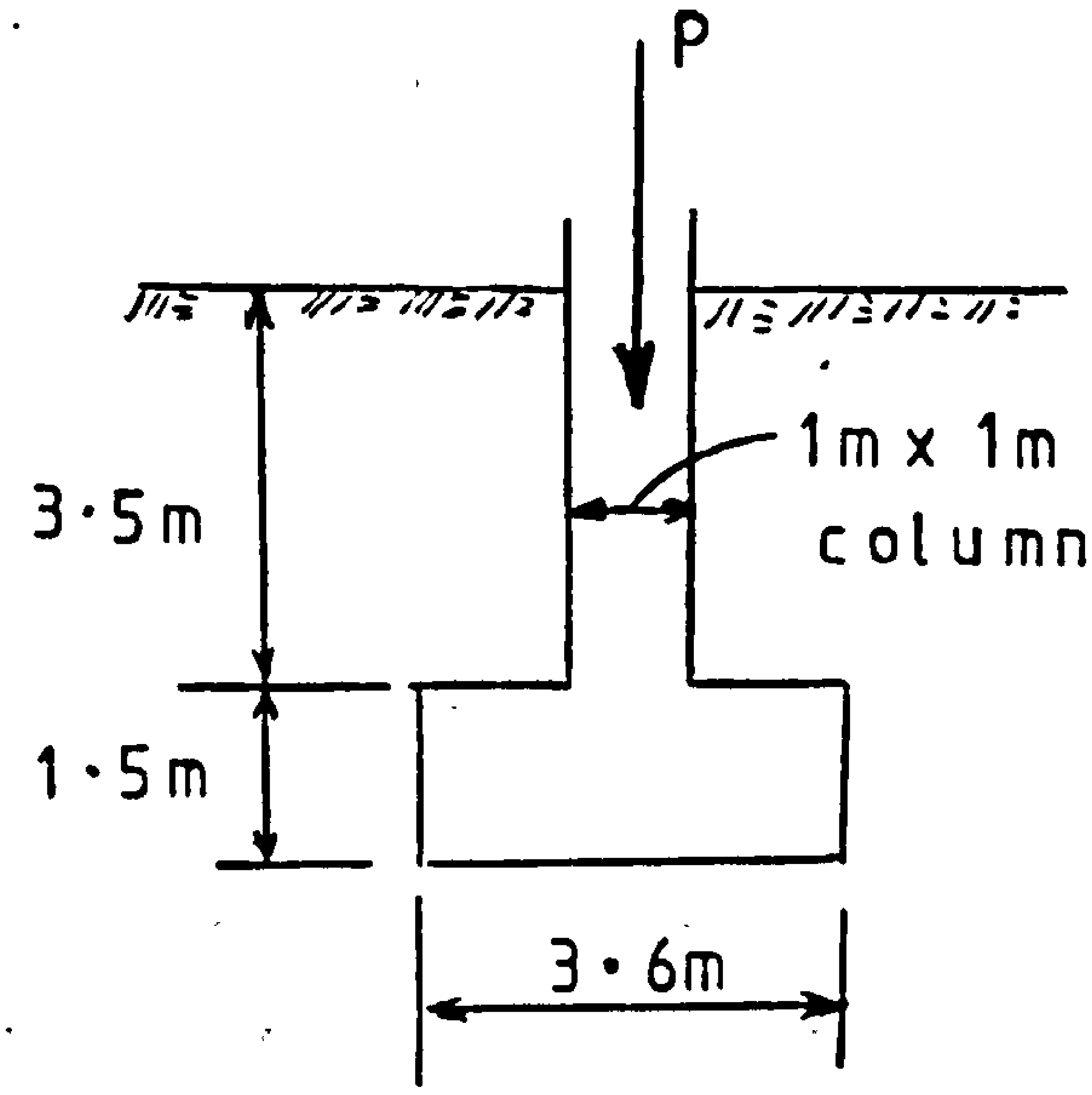


Fig.3.1

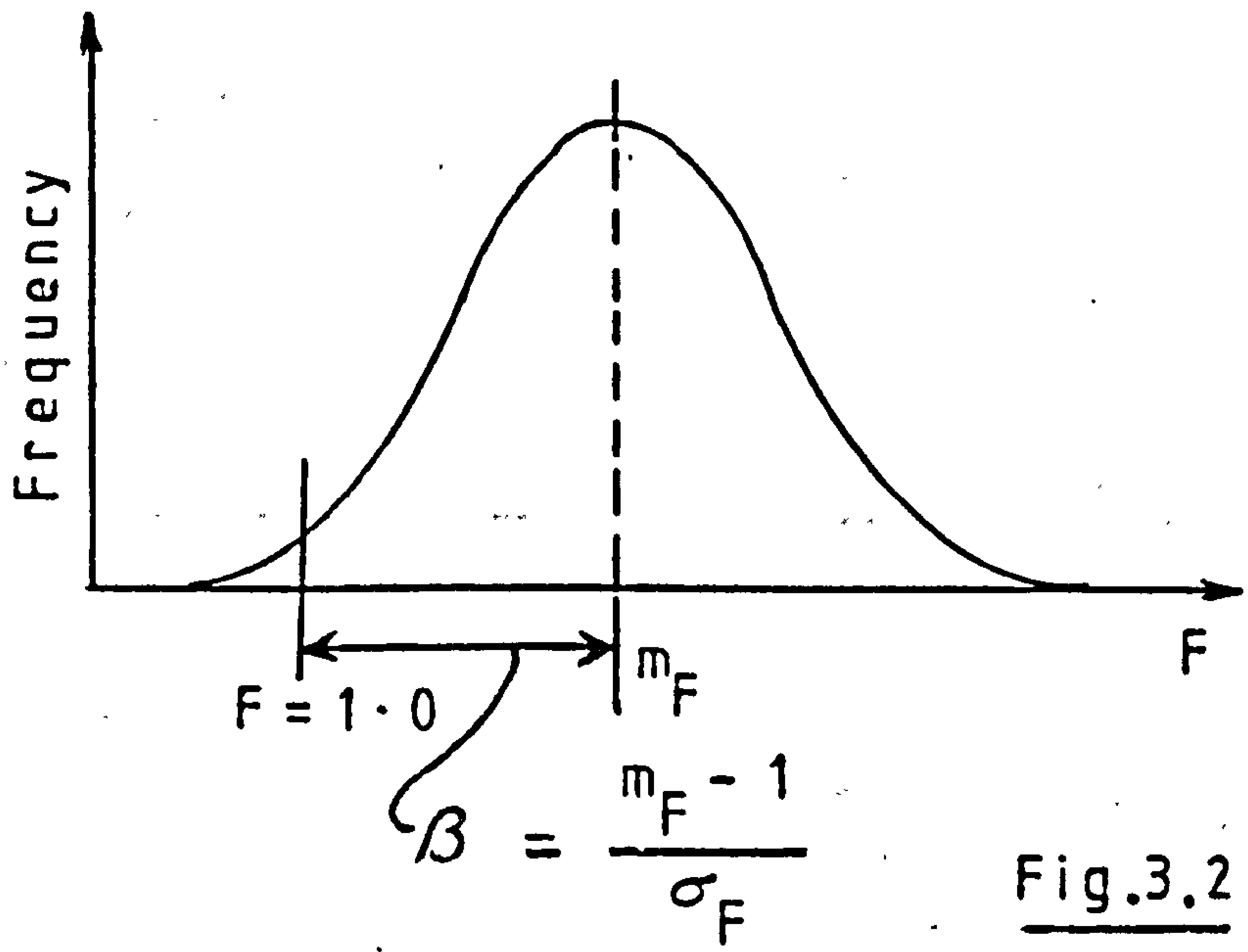


Fig.3.2

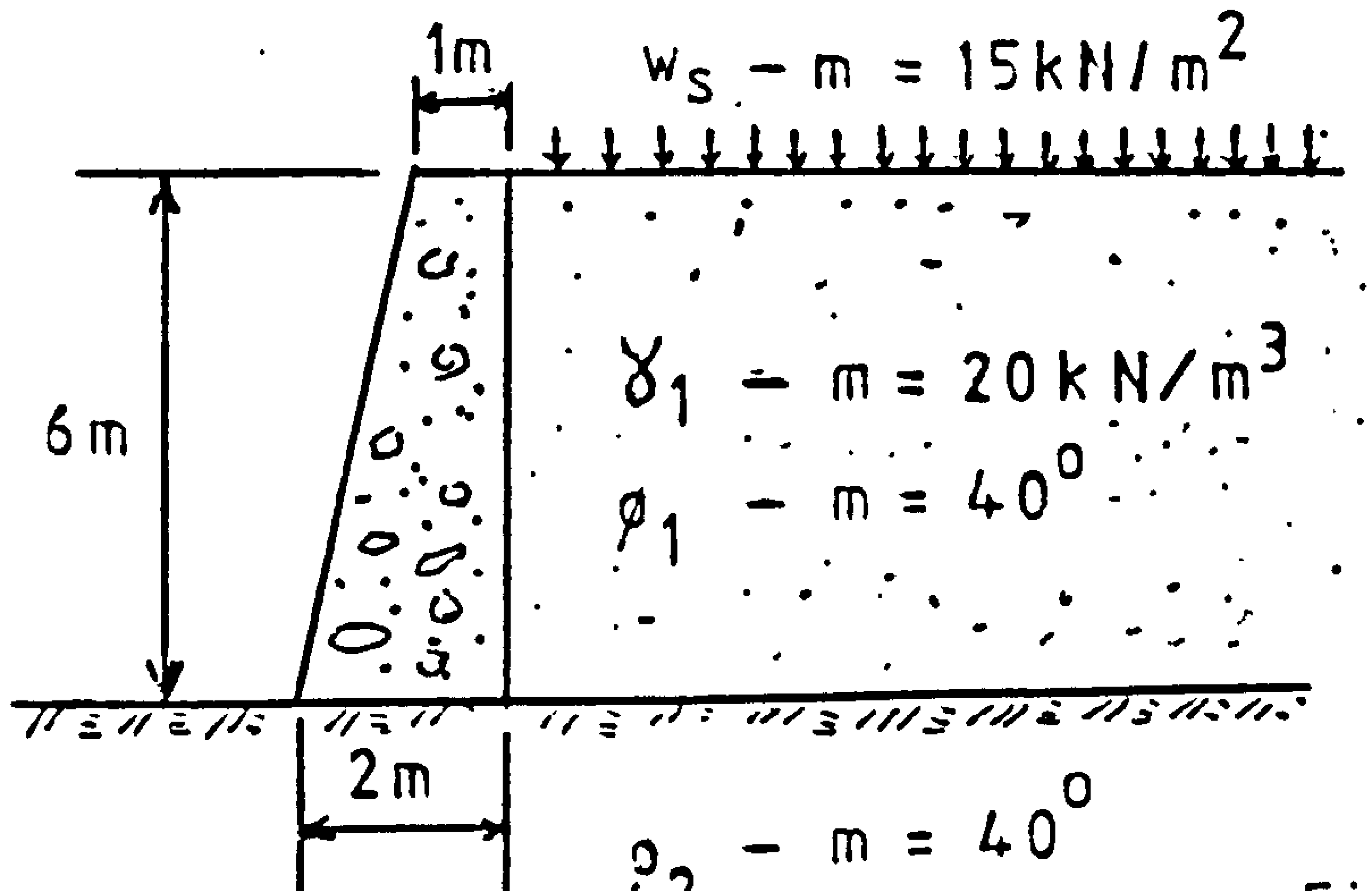


Fig.3.3

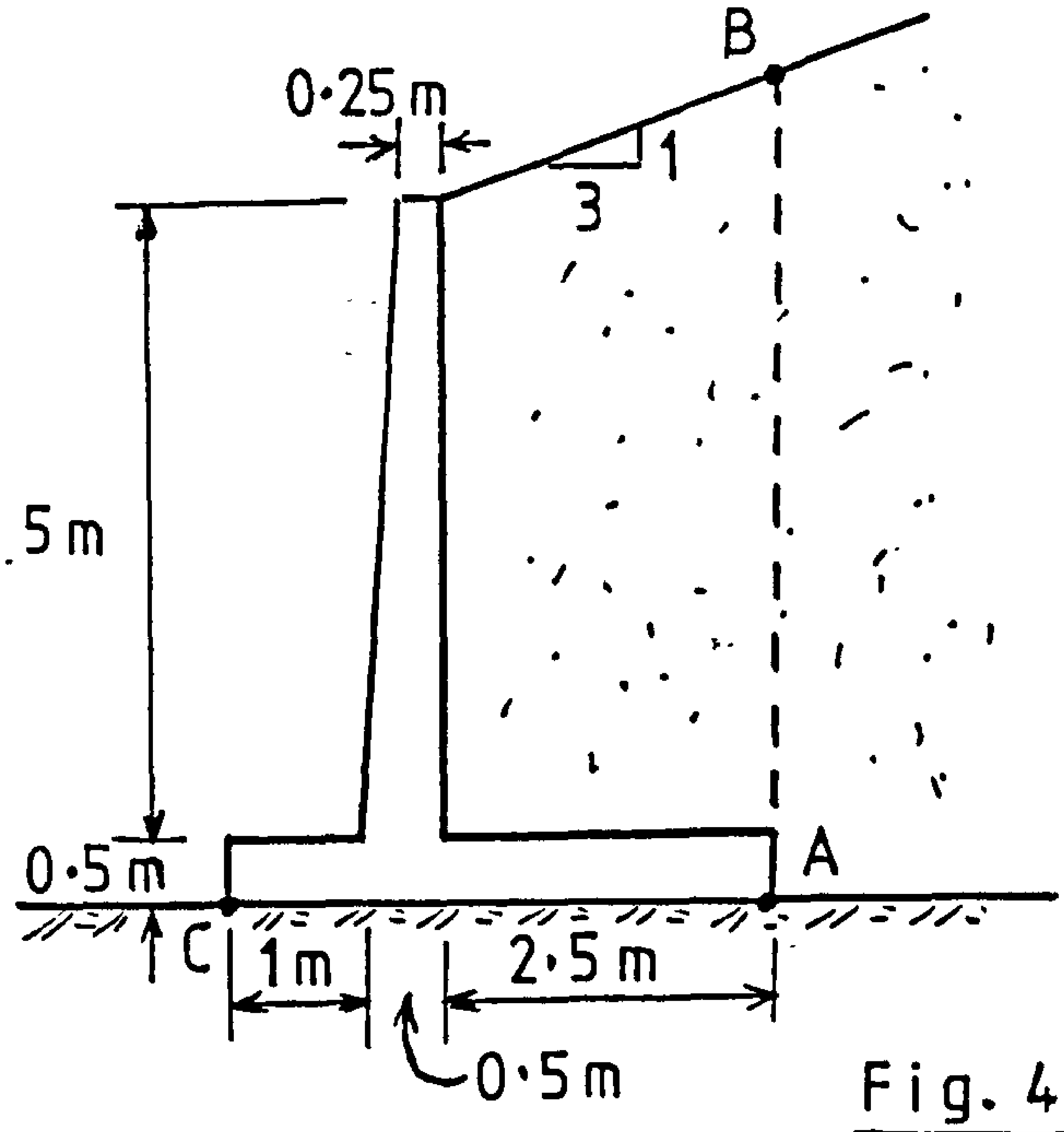


Fig. 4.1 Example 4.1

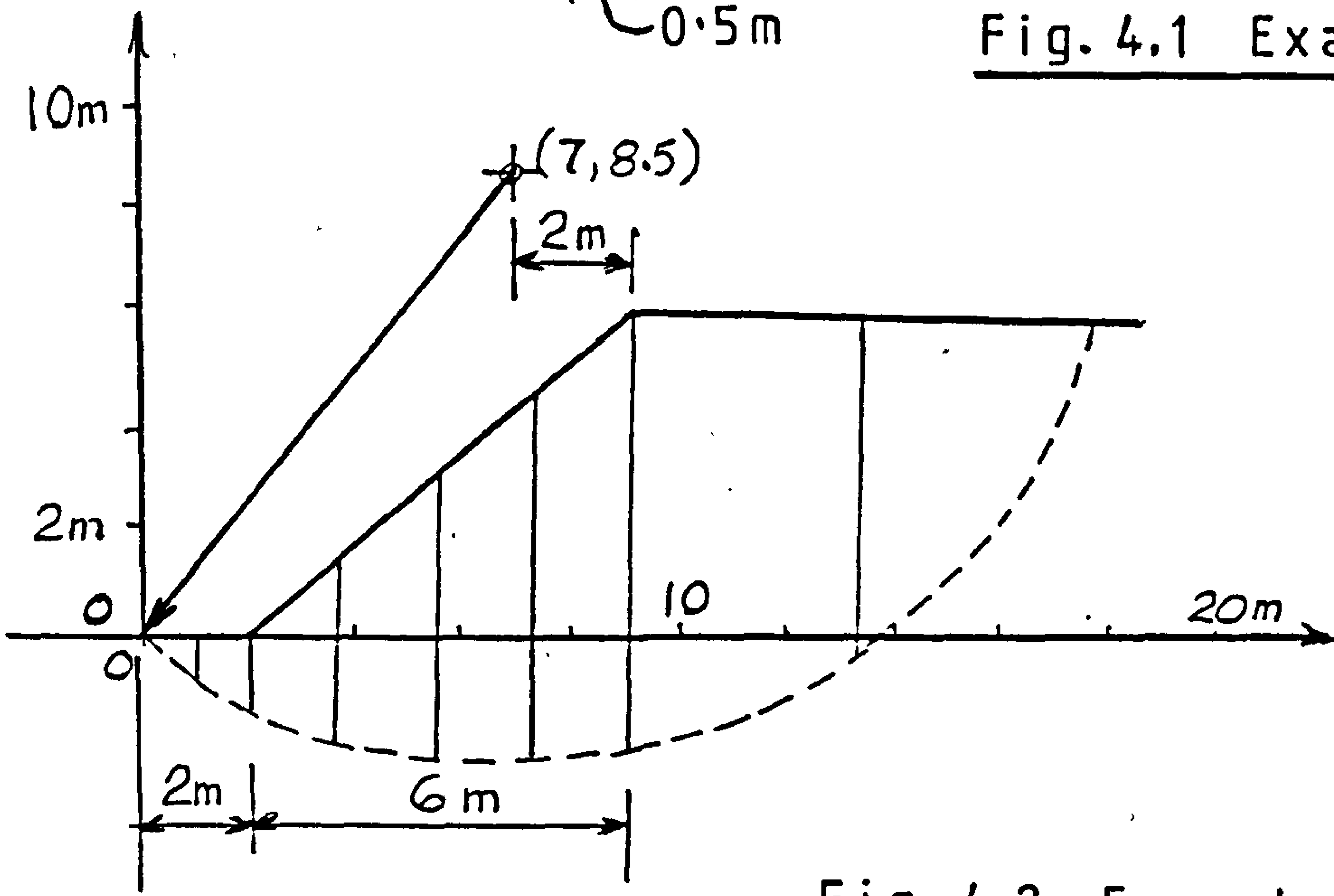


Fig. 4.2 Example 4.2

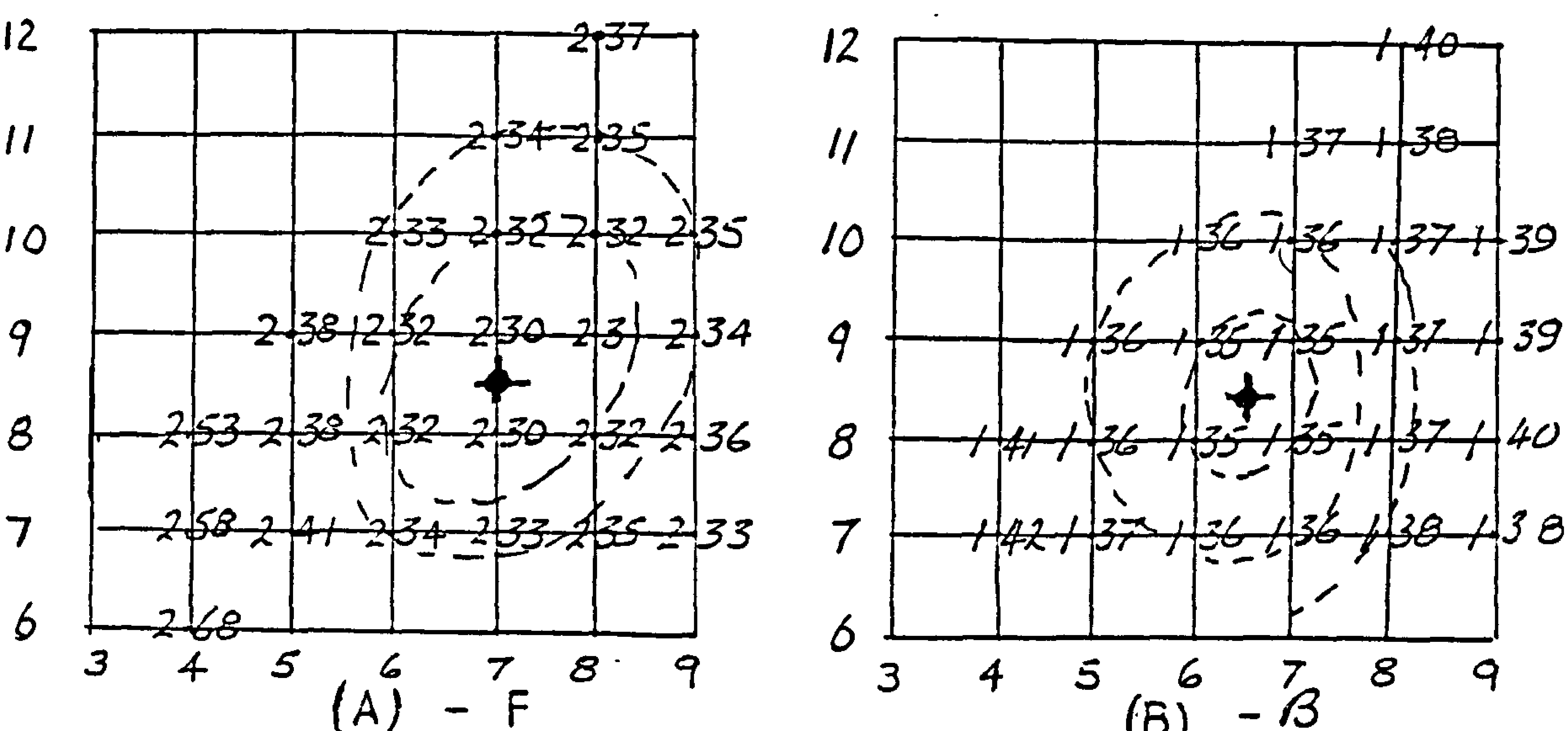


Fig. 4.3 Contours of F and β

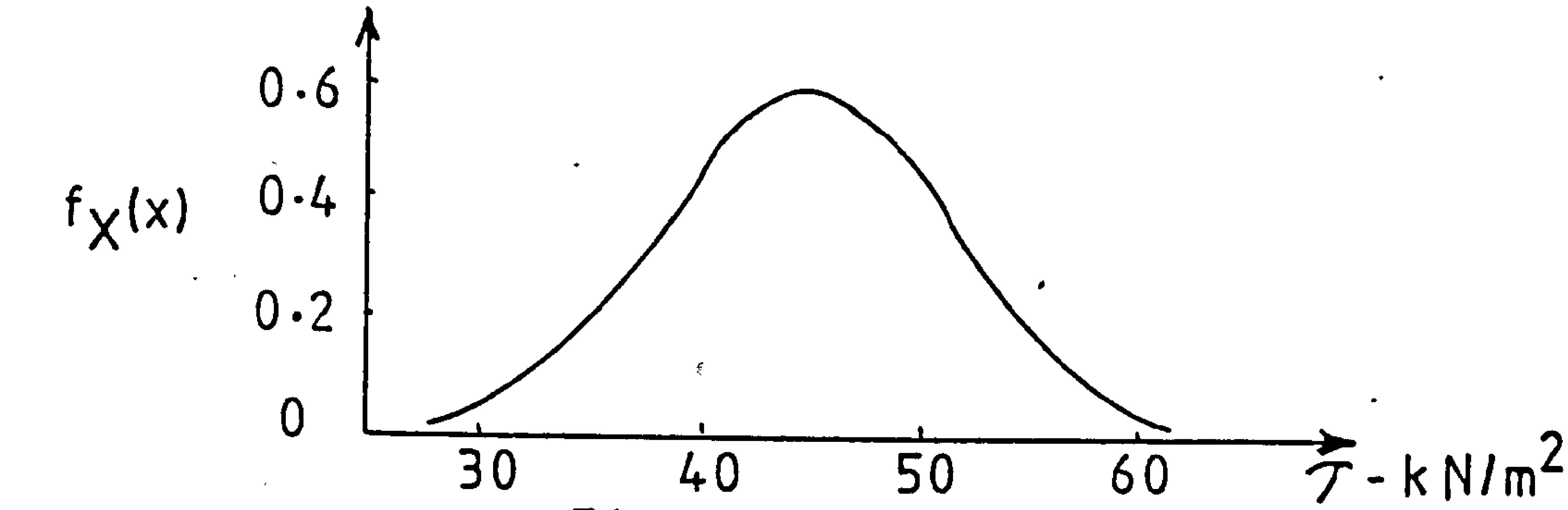


Fig. 5.1 Example 5.1

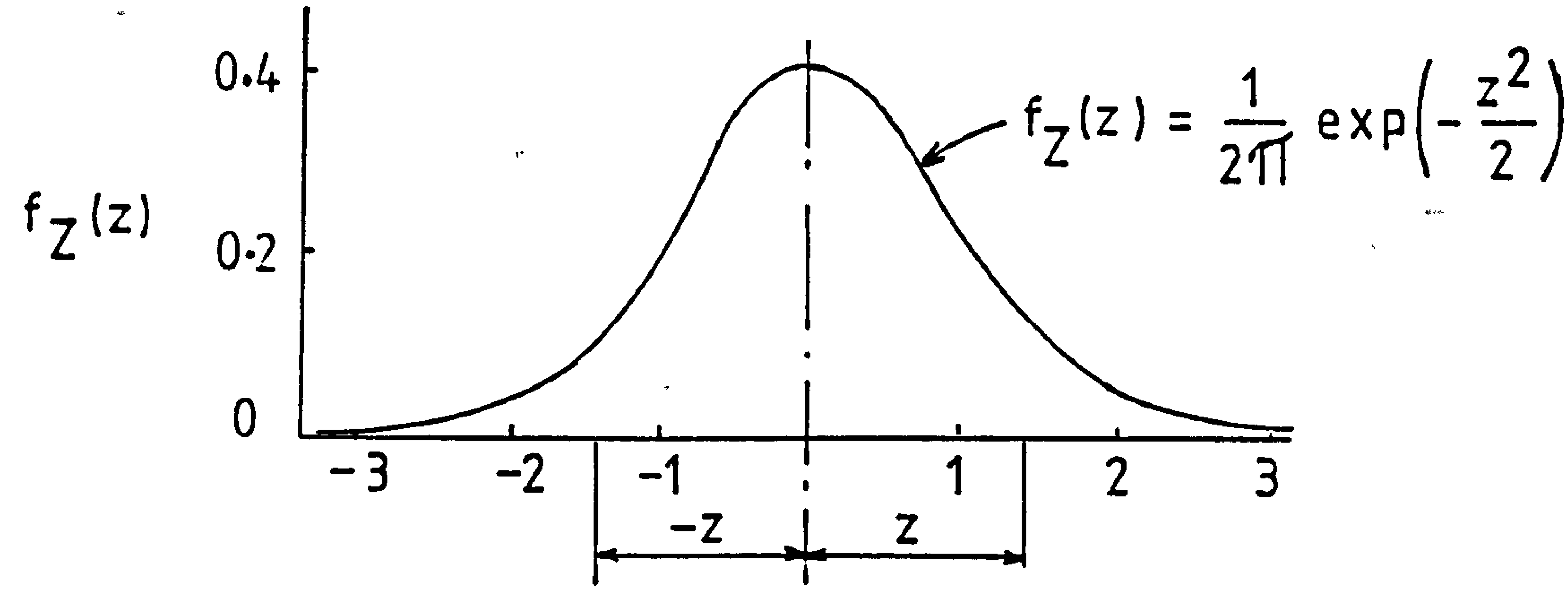


Fig 5.2 Example 5.2

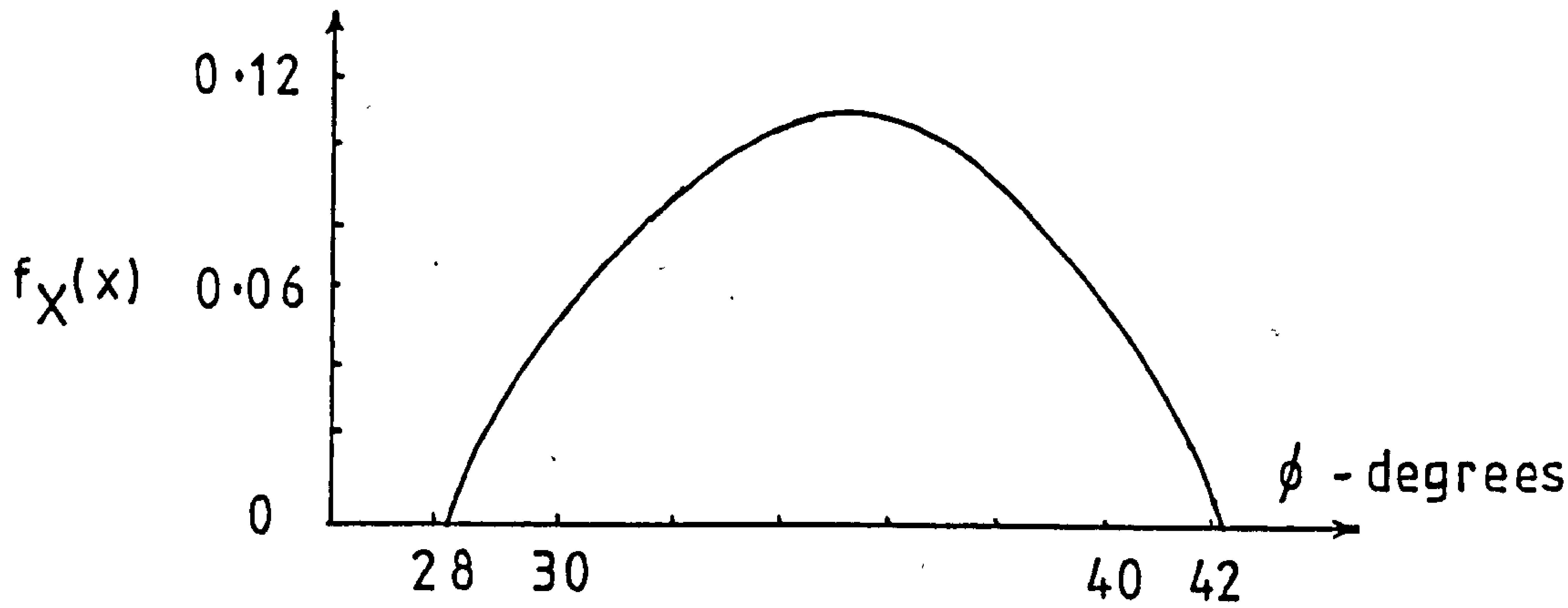


Fig. 5.3 Example 5.4

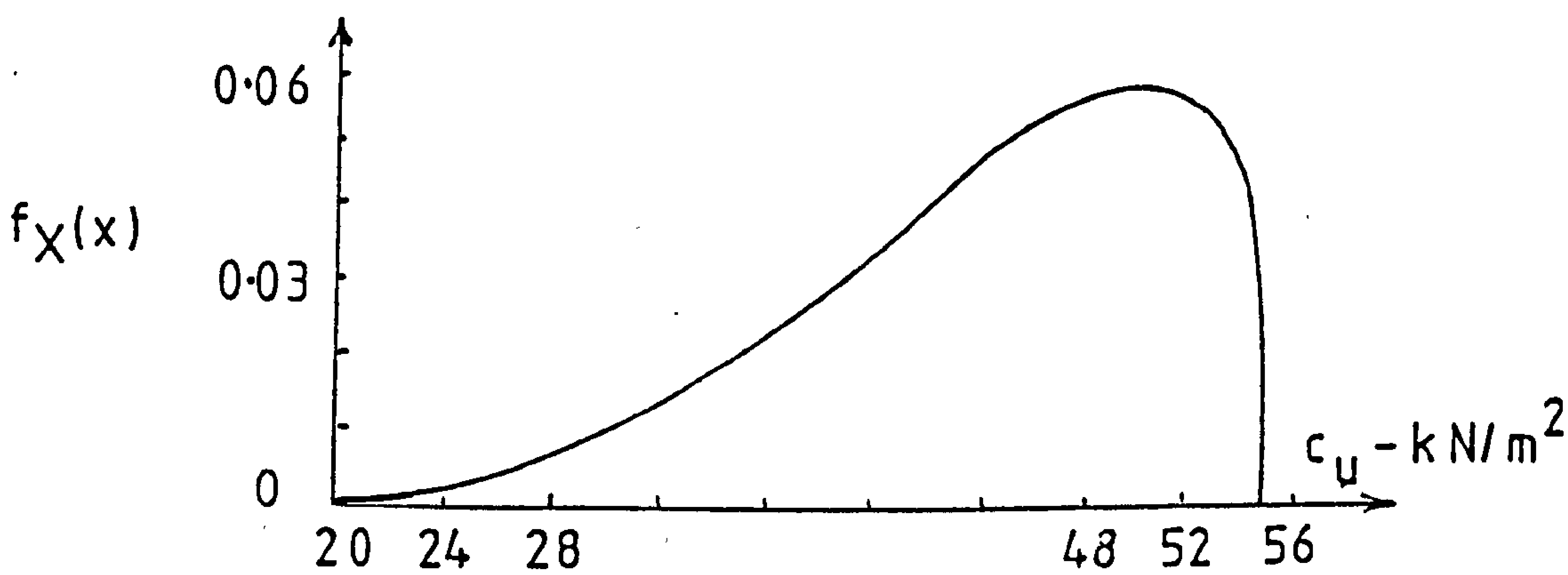


Fig. 5.4 Example 5.5

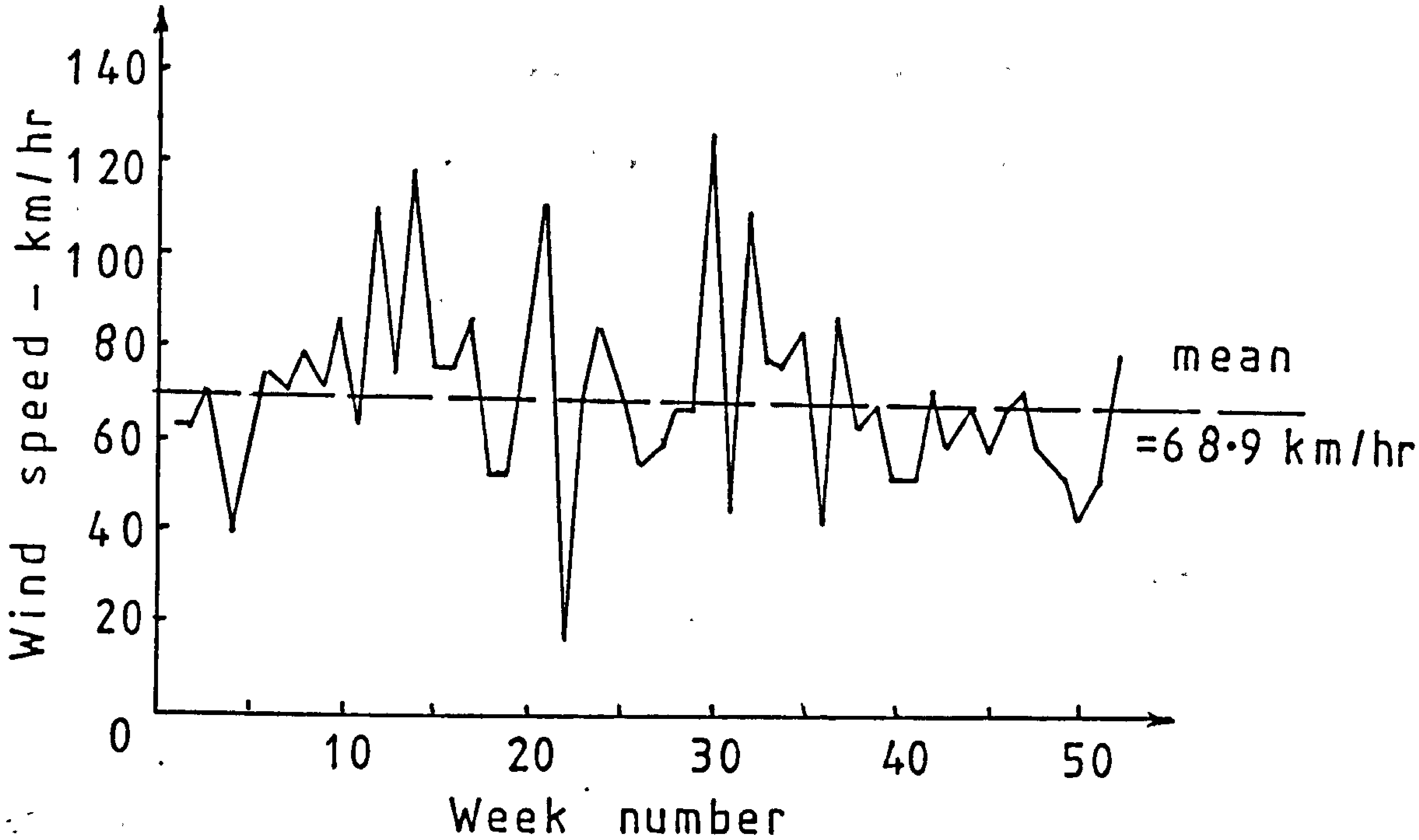


Fig. 6.1 Example 6.1 – Simulated wind speeds for one year

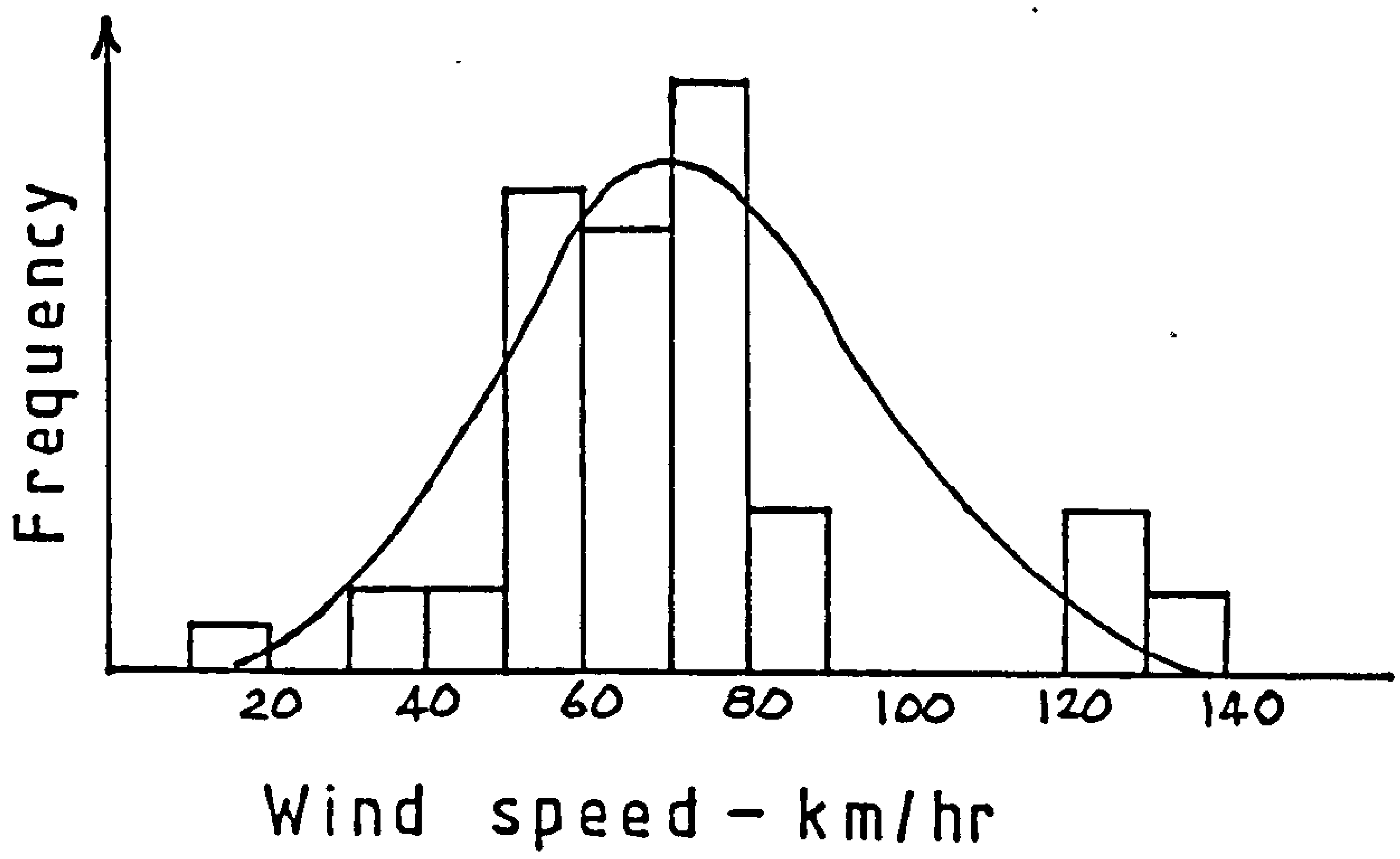


Fig. 6.2 Histogram of Fig. 6.1

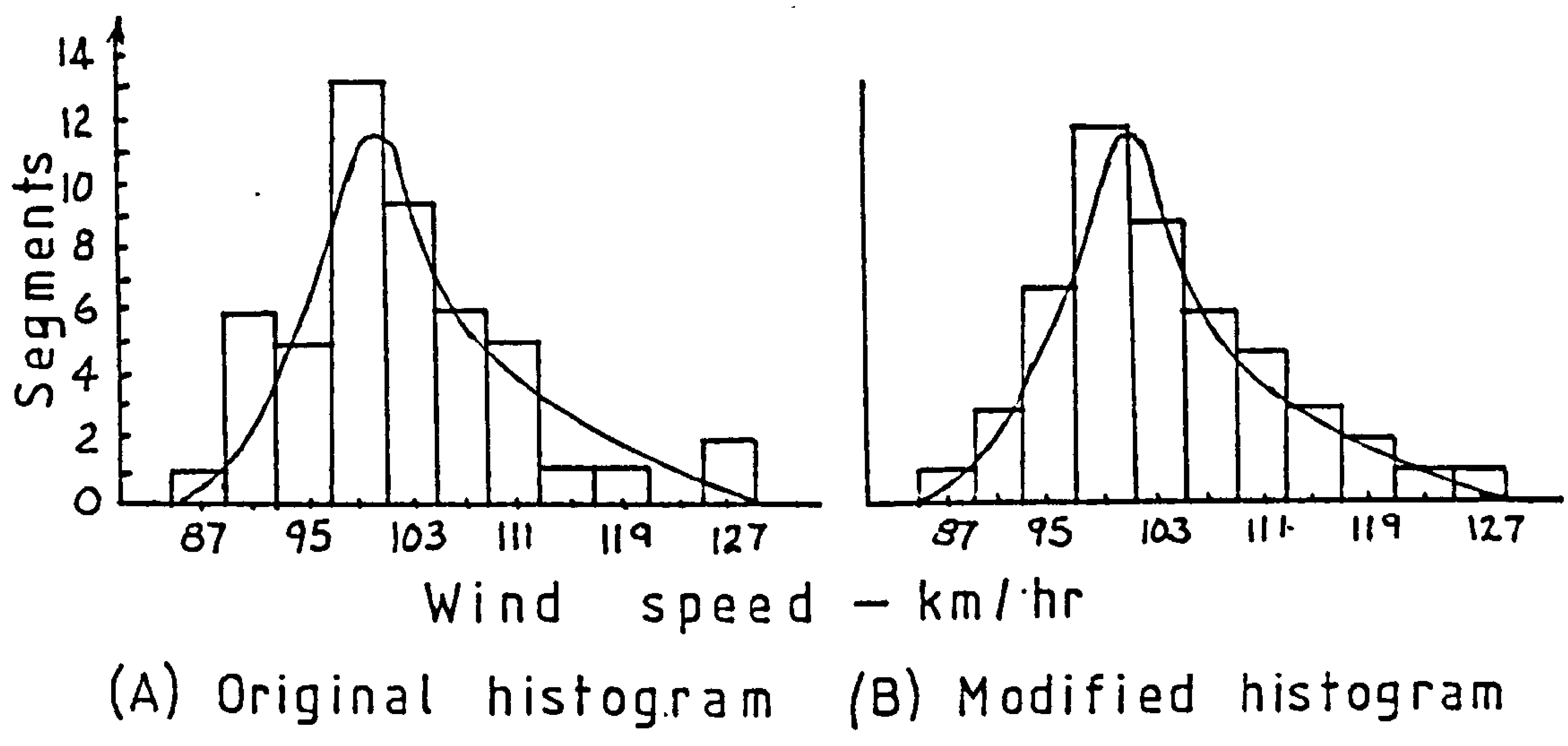


Fig. 6.3 Example 6.2

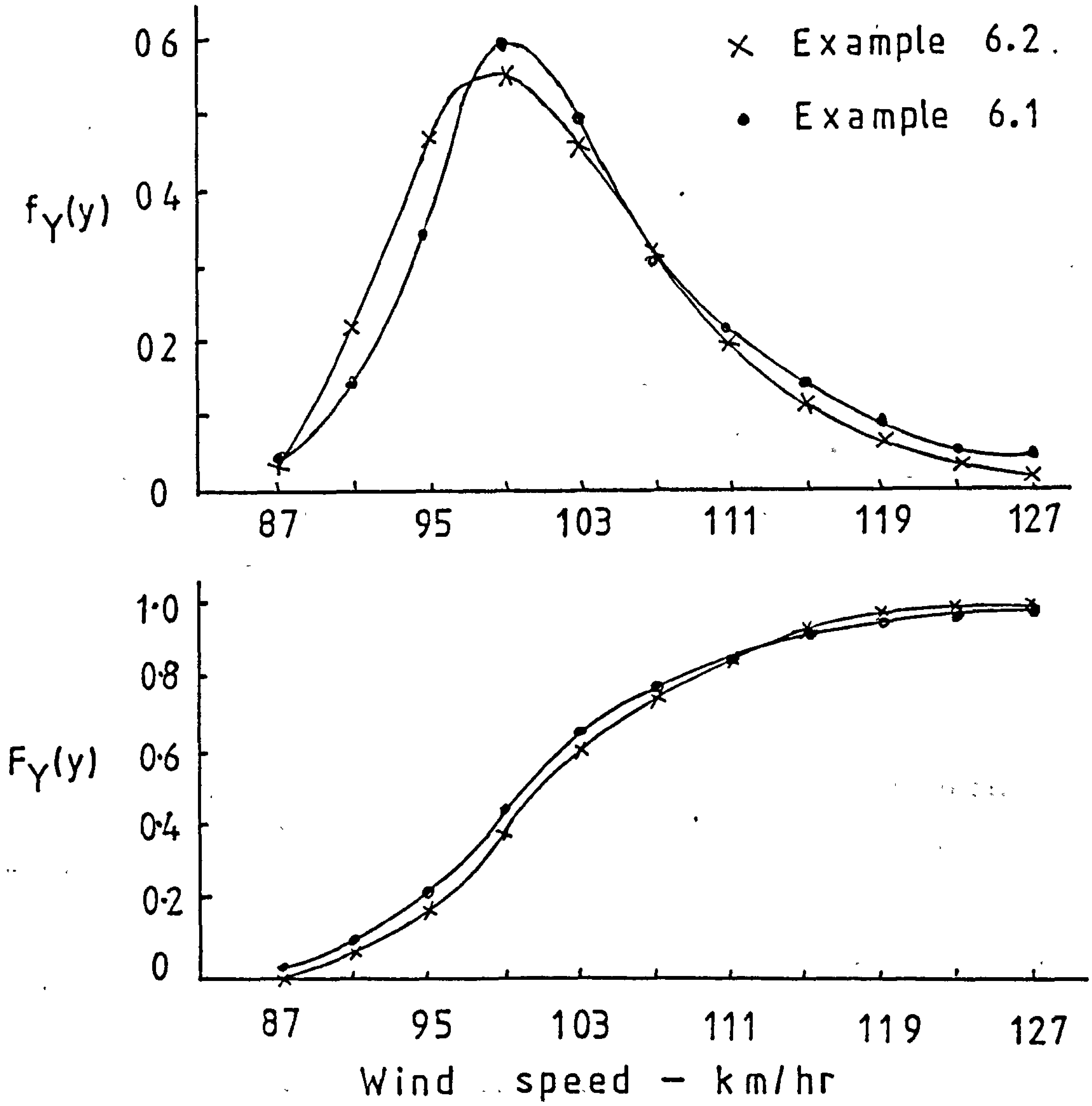


Fig. 6.4 pdf and cdf plots for Example 6.2

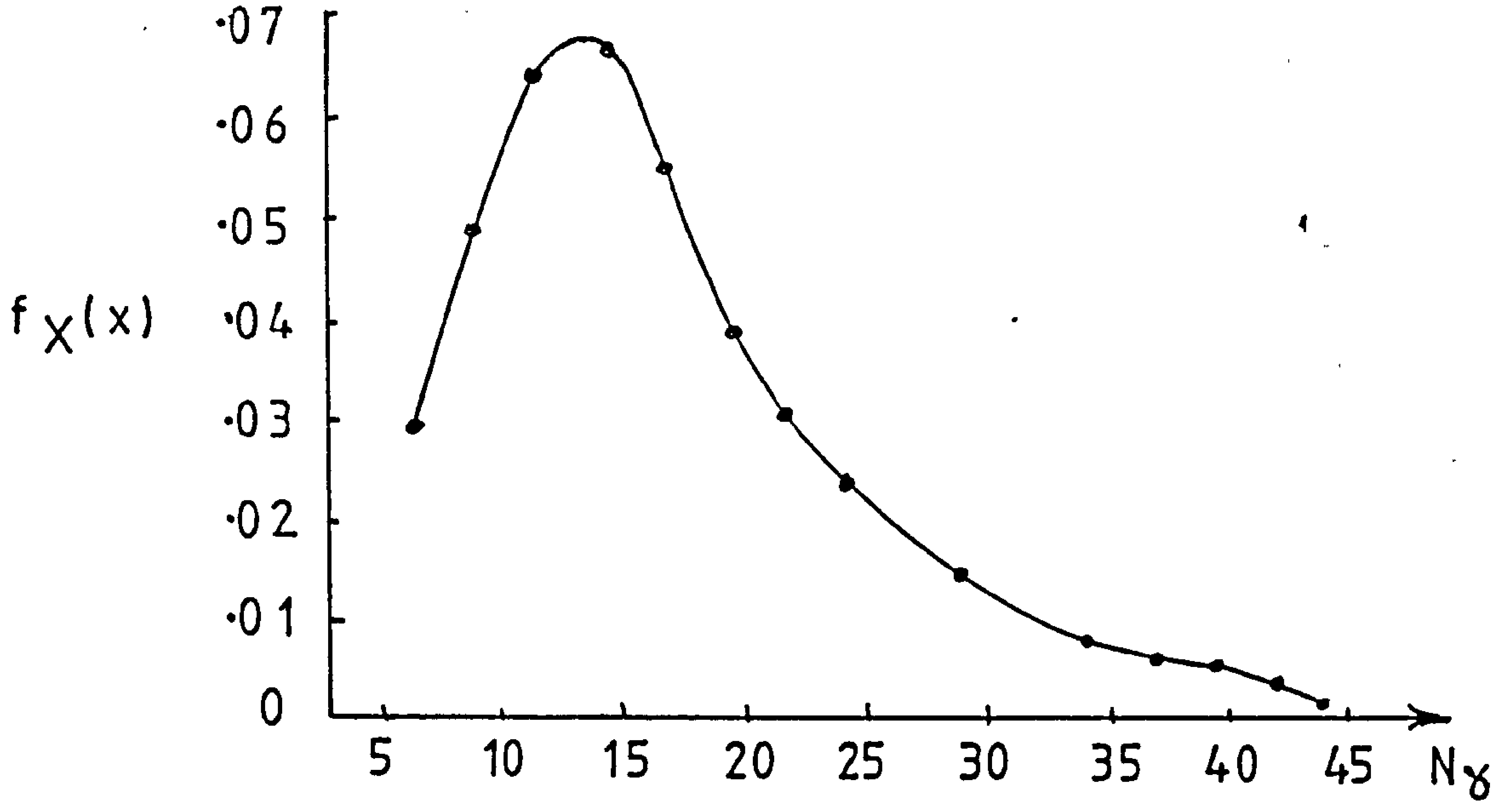


Fig. 7.1 Example 7.1 - Simulated pdf for N_γ

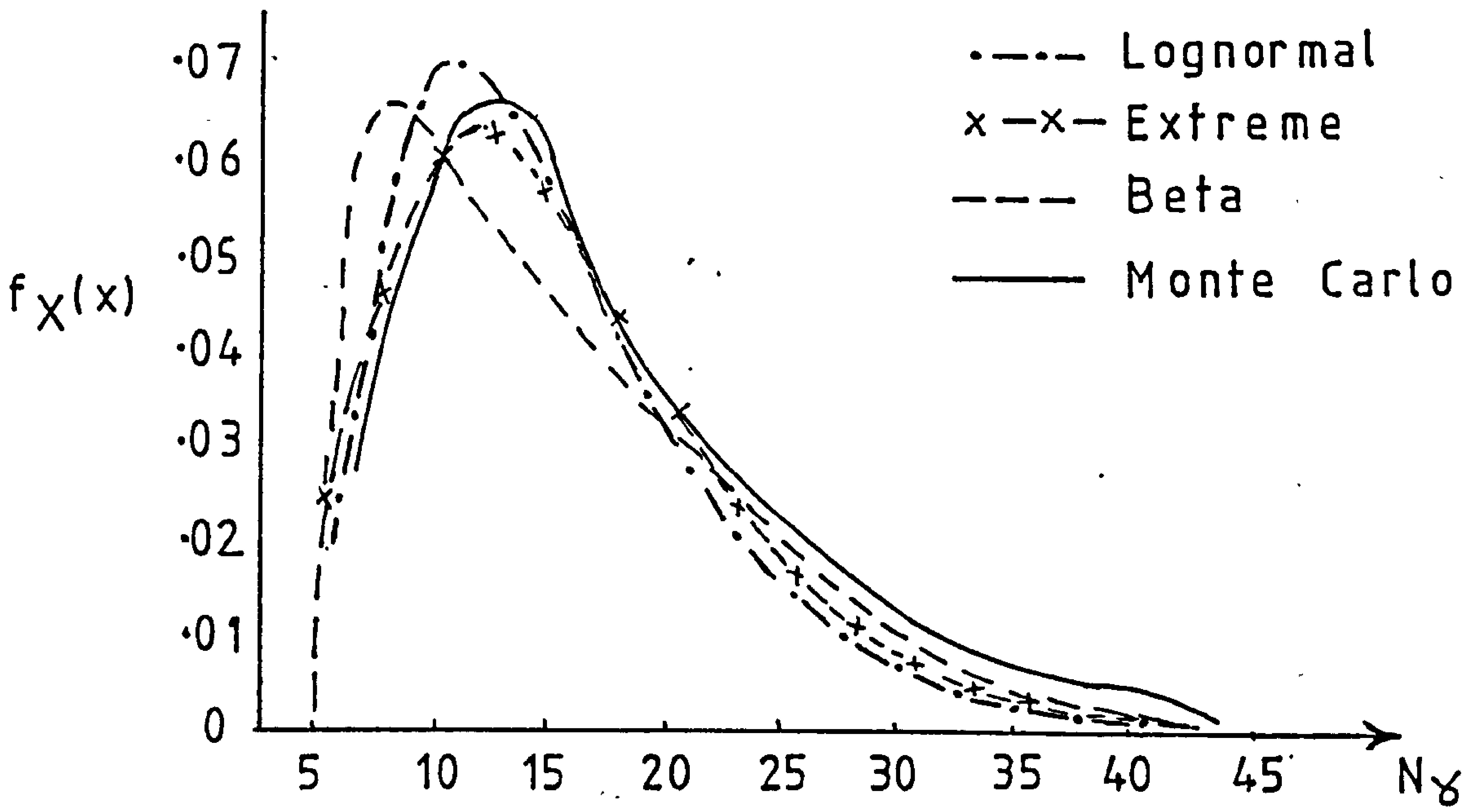


Fig. 7.2 Comparison of fitted distributions for N_γ

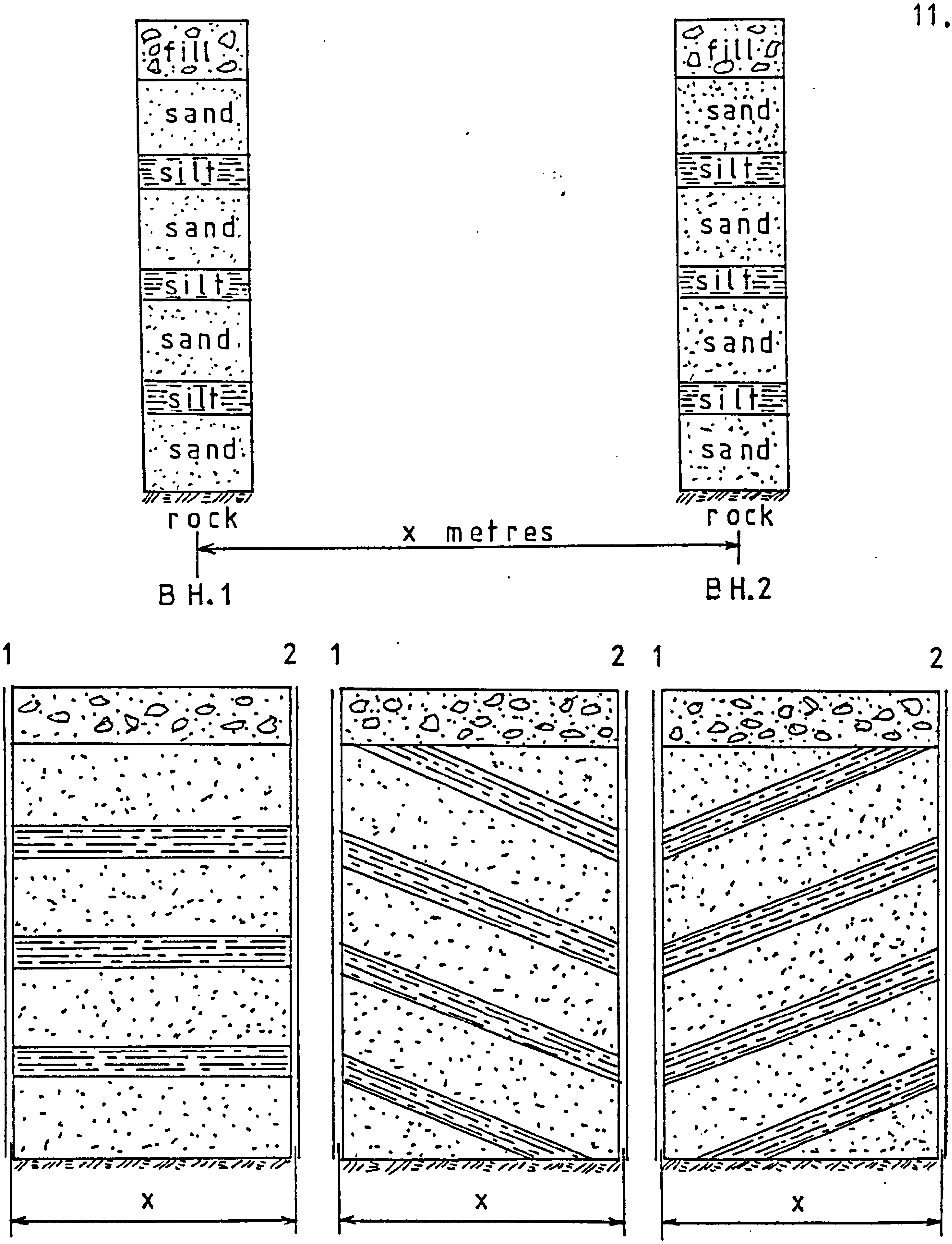


Fig. 8.1 Possible interpretations of similar borehole journals

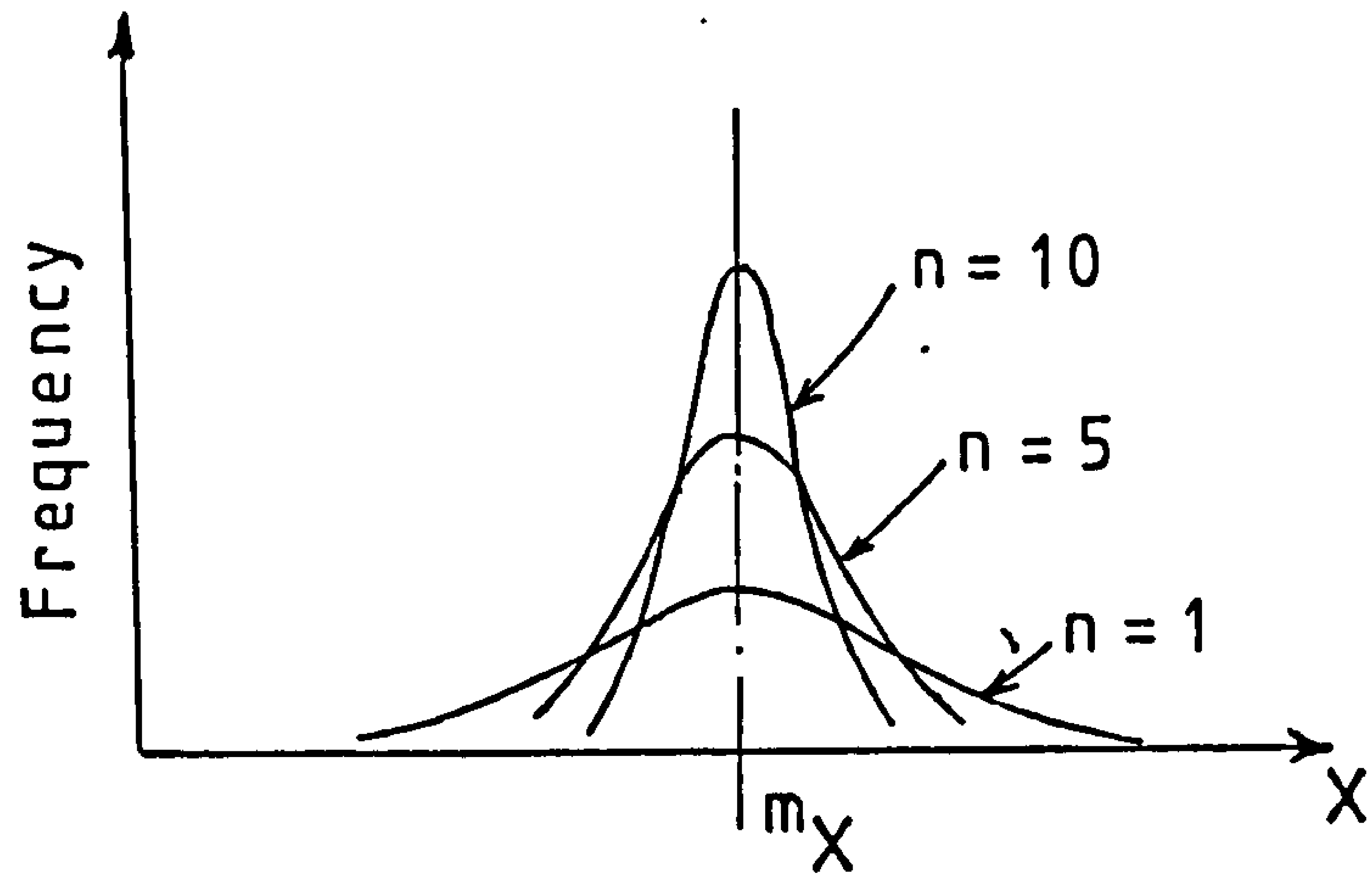


Fig.8.2 Frequency curves of mean values of groups

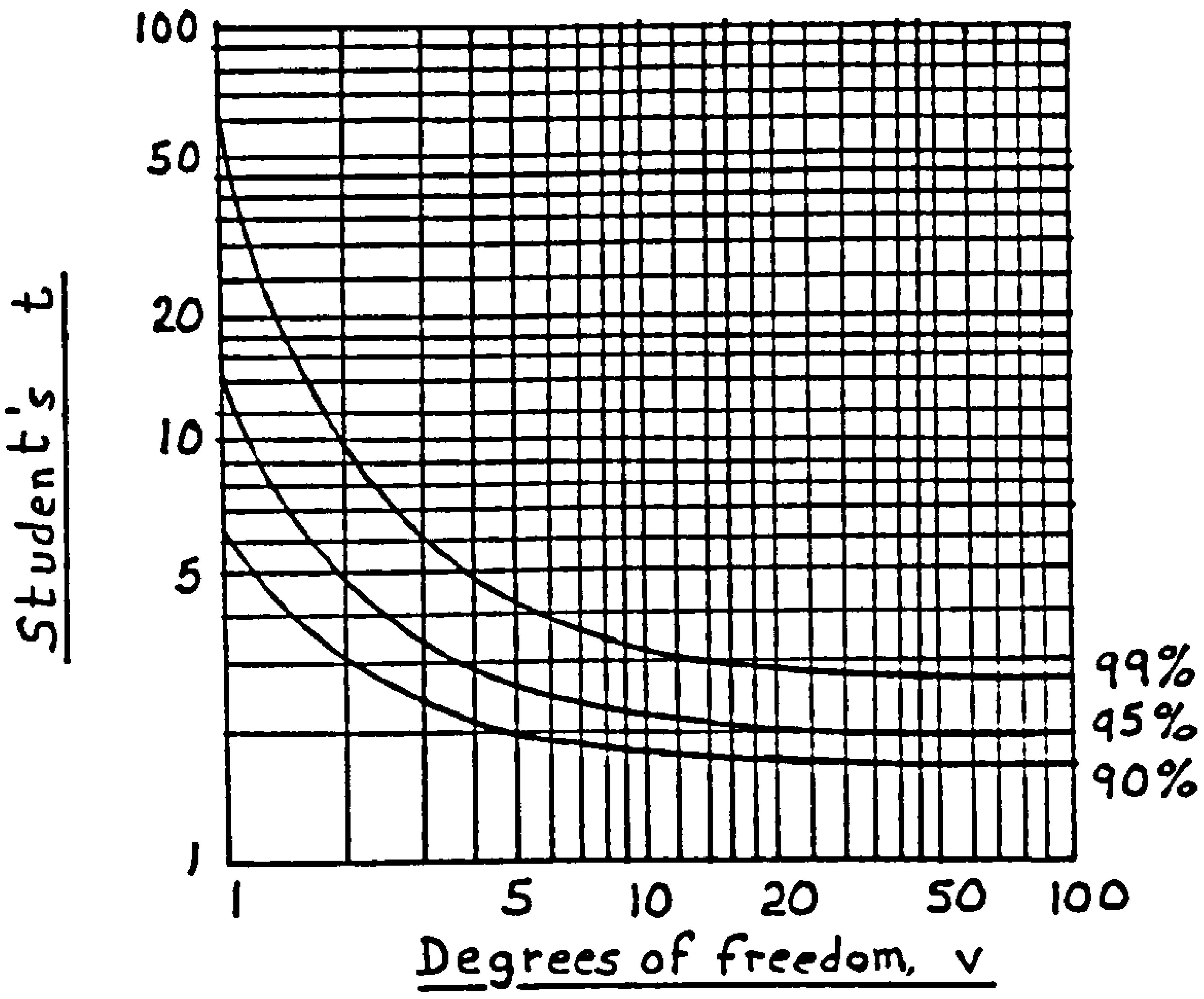


Fig. 8.3 Relationship between P, v and t

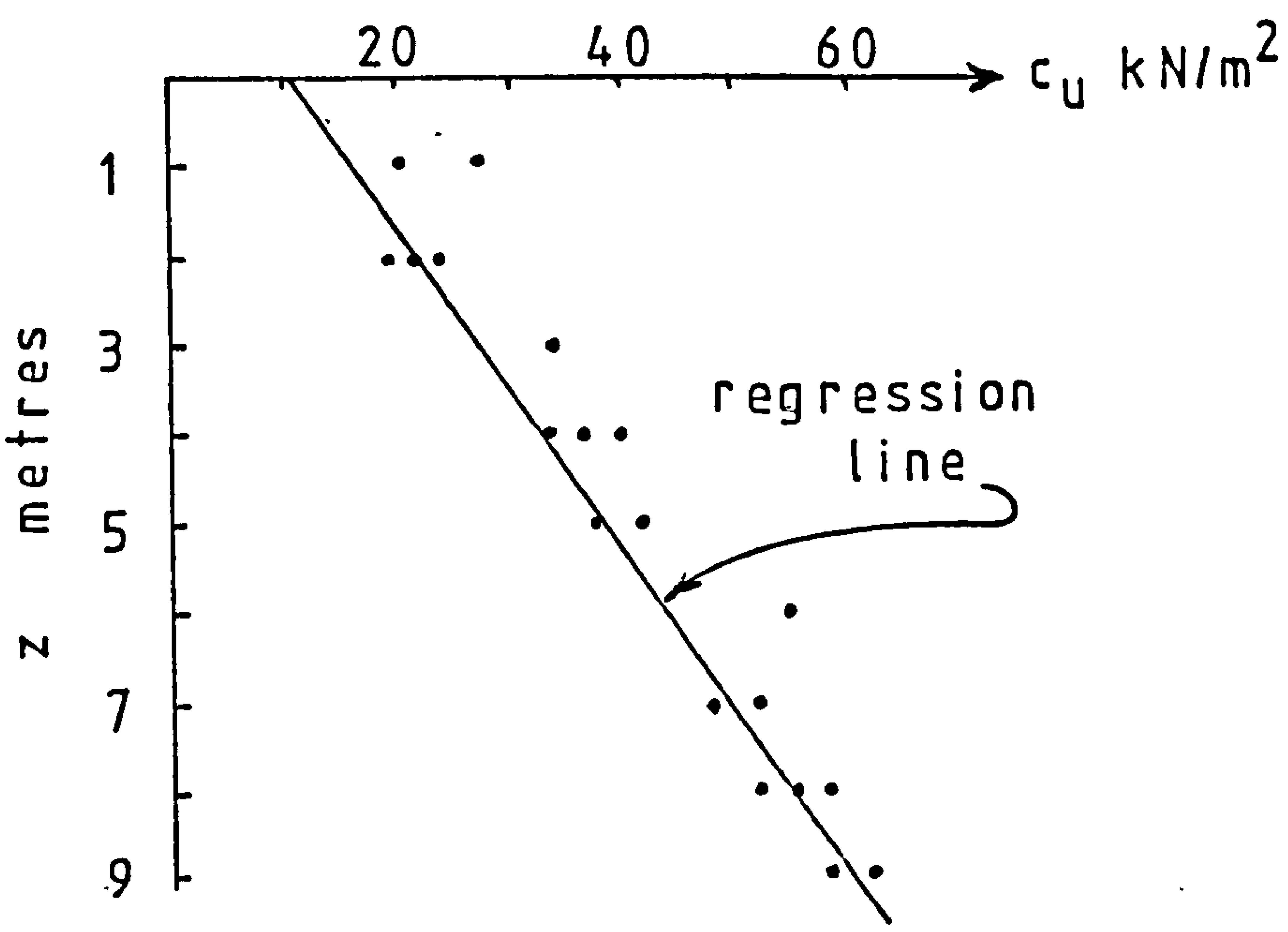


Fig. 8.4 Example 8.4

BIBLIOGRAPHY

ALONSON, E.E.
"Risk analysis of slopes and its application to slopes in Canadian sensitive clays",
Geotechnique, Vol.26, No.3, 1976.

BENJAMIN, J.R. & CORNELL, C.A.
" Probability, statistics and decison for civil engineers",
McGraw-Hill BoOK Co. 1970.

BODEN, J.B.
"Limit state principles in geotechnics",
Ground Engineering, Vol.14, No.6, 1981

BUTCHER, A.P.
"The current state of the ground investigation industry in the U.K.",
Ground Engineering, Vol.17 No.4, 1984.

CASAGRANDE, A.
"The role of the calculated risk in earthwork and foundation engineering", A.S.C.E. S.M.4, (1965)

CONSTRUCTION INDUSTRY RESEARCH & INFORMATION ASSOCIATION
"Rationalisation of safety and serviceability factors in structural codes",
Report No.63, 1976

CORNELL, C.A.
"Engineering seismic risk analysis",
Bulletin of Seismological Soc. of America, Vol. 58 No.5, 1968.

"Structural safety specifications based on second-moment reliability analysis",
I.A.B.S.E. Symposium on concepts of safety of structures and methods of design, London, (1969)

DE MELLO, V.F.B.
"Reflections on design decisions of practical significance to embankment dams",
Geotechnique, Vol.27 No.3 (1977)

DER KIEUREGHIAN, A. & ANG, A, H-S.
"Fault-rupture model for seismic risk analysis",
Bulletin of Seismologial Soc. of America, Vo. 67, No.4, 1977

DER KIUREGHIAN, A. & TAYLOR, R.L.
"Numerical methods in structural reliability",
Fourth Int. Conf. on application of statistics and probability in soil and structural engineering, 1983.

DITLEVESEN, O.
"Structural reliability and the invariance problem",
Solid Mechanics Division, University of Waterloo (Ontario),
Research Report 22, 1973.

"Evaluation of the effect on structural reliability of slight
deviations from hyperplane limit state surfaces",
D.I.A.L.O.G. 2-76, Danmarks Ingeniorakademi, Lyngby, (1976)

"Basic reliability concepts",
N.A.T.O. Advanced Study Institute on reliability theory and its
application in structural and soil mechanics, Bornholm, Denmark,(1982)

DRISCOLL, R.
"Eurocode 7 - Foundations",
Ground Engineering, Vol.17, No.1, 1984

FERRY BORGES, J. & CASTANHETA, M.
"Structural safety",
National Laboratory of Civil Engineering (Lisbon), 1974.

FIESSLER, B.
"Das programmsystem FORM zur berechnung der
versagenswahrscheinlichkeit vo komponenten von tragsystmen",
Sonderforschungsbereich 96, Technical University of Munich, 1980.

FIESSLER, B. & RACKWITZ, R.
"Note on discrete safty checking when using non-normal stochastic
models for basic variables",
Load project working session, M.I.T.,Cambridge, MA. 1976.

FIESSLER, B., NEUMANN, H.-J., RACKWITZ, R.
"Quadratic limit states in structural reliability",
Jour. Eng. Mech. Div., A.S.C.E., Vol. 105, 1979.

FISHER, R.A. & TIPPETT, L.H.C.
"Limiting forms of the frequency distribution of the largest or
smallest member of a sample",
Proc. Cambridge Philosophical Society, Vol.24, 1928.

GRIGORIU, M.
"Risk analysis for code calibration",
Structural Safety, Vol.1, No.4, Elsevier Science Publishers B.V.,
Amsterdam, 1983.

GUMBEL, E.J.
"Statistics of extremes",
Columbia University Press, 1958.

HAMMITT, G.M.
"Statistical analysis of data from a comparative laboratorytest
program sponsored by A.C.I.L.",
U.S. Army Eng. Waterways Exp. Station, Corps of Engineers Misc.,
Paper no. 4-758, 1966.

HANSEN, J.B.
"A revised and extended formula for bearing capacity",
Institut Kobenhavn Bulletin, 1970.

HASOFER, A.M. & LIND, N.C.
"Exact and invariant second-moment code format",
Jour. Eng. Mechs. Div., A.S.C.E., Vol.100, 1974.

HOOPER, J.A. & BUTLER, F.G.
"Some numerical results concerning the shear strength of London clay",
Geotechnique, Vol.16, No.4, 1966.

HOHENBICHLER, M. & RACKWITZ, R.
"Non-normal dependent vectors in structural reliability",
Jour. Eng. Mech. Div., A.S.C.E., Vol. 107, 1981.

KITAGAWA, M. & DER KIUREGHIAN, A.
"Safety index by first-order second-moment reliability method",
Report No. UCB/SESM-80/90, Div.Struct. Eng. & Mechs., Univ. of California, Berkly, 1980.

LADD, C.C.
"Stress deformation and strength characteristics," Proc. 9th I.C.S.M.F.E., Tokyo (1977)

LUMB, P.
"The residual soils of Hong Kong",
Geotechnique, Vol.15, No.2, 1965.

"The variablity of natural soils",
Canadian Geotechnical Jour., Vol.3, No.2, 1966.

"Safety factors and the probability distribution of soil strength",
Canadian Geotechnical Jour., Vol.7, No.3, 1970.

"Applications of statistics in soil mechanics",
'Soil mechanics - New horizons', Pub. Butterworth, London, 1974.

MAYER, H.
"Die sicherheit der bauwerke",
Springer - Verlag (Berlin), 1926

MEYERHOF, G.G.
"Influence of roughness of base and groundwater conditions on the ultimate bearing capacity of foundations",
Geotechnique, Vol.5, No.2, 1955.

"Safety factors in soil mechanics",
Canadian Geotechnical Jour. Vol.7, No.4, 1970.

"Limit state designs in geotechnical engineering"
Structural safety, Vol.1 No.1, Elseviers Science Publishers B.V., Amsterdam, 1982.

MORSE, R.K.
"The importance of proper soil units for statistical analysis",
Statistics and Probability in CivilEngineering (ed. P. Lumb),
Hong Kong University Press, Hong Kong, 1971.

MOSTYN, G.
"A statistical approach to characteristing the permeability of a mass",
Fourth Int. Conf. on application of statistics and probability in soil and structural engineering, 1983.

OVESEN, N. K.
"Towards a European code for foundation engineering",
Ground Engineering, Vol.14, No.7, 1981.

PECK, R. B.
"Advantages and limitations of the observational method in applied soil mechanics"
Geotechnique Vol.19 No.2 (1969).

RACKWITZ, R.
"Practial probabalistic approach to design",
Technical University of Munich, 1976.

RAVINDRA, M.K., HEANEY, A.C. & LIND, N.C.
"Probabilistic evaluation of safety factors",
I.A.B.S.E. Symposium on concepts of safety of structures and methods of design, London, 1969.

ROSENBLUETH, E. & ESTEVA, L.
"Reliability basis for some Mexican codes",
American Concrete Institute Publication SP-31, 1972

SAETRE, H.J.
"On high wave conditions in the Northern North Sea",
Oceanology International 75, Conference papers, Brighton, U.K.,1975

SCHULTZE, E.
"Frequency distributions and correlations of soil properties",
First Int. Conf. on applications of statistics and probabiliy in soil and structural engineering, 1972

SHERWOOD,P.T.
"The reproducibity of the reults of soil classification and compaction tests",
Road research Laboratory Report, LR339, 1970.

SIMPSON, B.
"Eurocode EC7 - the preparation of a base document",
Ground Engineering, Vol.16, No.3, 1983

SMITH, G.N.
"An introduction to matrix and finite element methods in civil engineering",
Applied Science Publishers Ltd., London, 1971

"Elements of soil mechanics for civil and mining engineers",
5th. Edition, Granada Pub. Ltd., London, 1982.

"Probability theory in geotechnics - an introduction",
Ground Engineering, Vol.14, No.7, 1981

SMITH, G.N. & POLE, E.L.
"Elements of foundation design",
Granada Pub.Ltd., London, 1980.

SOKOLOVSKII, V.V.
"Statics of soil media",
Butterworth, London, 1960.

TANG, W.H.
"Probabilistic evaluation of loads",
Jour. Eng. Mech. Div., A.S.C.E., Vol. 107, 1981.

TURKSTRA, C.J.
"Theory of structural design decisions",
University of Waterloo (Ontario), Solid Mechanics Study 2, 1970.

TURNBULL, W.J., COMPTON,J.R. & AHLVIN, R.G.
"Quality control of compacted earthwork",
Jour. Soil Mechs. and Found. Div., 1966

UNO,T., ARAI,H. & SHIBAYAMA,M.
"Models for predicting ground water level",
I.C.S.M.F.E., Stockholm (1981)

YUCEMAN, M.S. & TANG, W.H.
"Long term stability of slopes - a reliability approach",
Second Int. Conf. on application of statistics and probability in soil and structural engineering, 1975.
